

Memòria del projecte

Construcció i explotació d'un magatzem de
dades de planificació hidrològica

A la Sònia per la seva paciència infinita durant aquests anys de carrera.

Agraïments

A tots els professors i consultors de la UOC que m'han ajudat durant aquests anys de carrera.
Als pirates de Silicon Valley per la seva inspiració.

Resum

Aquest projecte neix amb la finalitat d'ajudar al nostre client, la Confederació Hidrogràfica Nord-est (CHNE), a ser més eficient en el tractament de la seva informació. La CHNE és l'organisme que desenvolupa tasques de planificació hidrològica, gestiona l'aigua i construeix les infraestructures hidràuliques al nord-est de la península ibèrica.

Aquest increment d'eficiència s'aconseguirà mitjançant la implementació d'un Magatzem de Dades, que facilitarà als tècnics del departament de planificació hidrològica, les tasques de recollida i explotació de la informació.

La primera part del projecte consisteix en la definició tant de les Dimensions i Atributs com dels Indicadors del model de dades i per això caldrà: analitzar en detall les dades proporcionades pel nostre client i les seves peticions d'informació.

A continuació es dissenya i es crea un Magatzem de Dades fent servir Oracle Express. Aquest magatzem de dades s'alimentarà amb les dades prèviament transformades, per adaptar-les a l'estructura necessària per la seva càrrega (Procés ETL).

Per últim es creen els diferents informes i gràfics. Aquesta part del projecte és la que realment aporta un valor afegit al nostre client doncs l'ajuda en el procés de presa de decisions.

Paraules clau

Magatzem de Dades (DW), ETL, SQL, Confederació Hidrogràfica Nord-Est (CHNE), Business Intelligence (BI), Oracle Discoverer, Oracle Express.

Índex de continguts

1. INTRODUCCIÓ	
1.1 Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC.....	4
1.2 Objectius del TFC.....	4
1.2.1 Vessant d'estudiant d'ETIG.....	4
1.2.2 Vessant Client.....	5
1.3 Enfocament i Mètode Seguit.....	5
1.4 Planificació del Projecte.....	5
1.4.1 Tasques prèvies.....	5
1.4.2 Realització PAC1.....	6
1.4.3 Realització PAC2.....	6
1.4.4 Realització PAC3.....	6
1.4.5 Lliurament final.....	6
1.4.6 Debat Virtual final.....	7
1.4.7 Diagrama de Gantt.....	7
1.5 Productes Obtinguts.....	9
1.6 Descripció dels altres capítols de la memòria.....	9
1.6.1 Anàlisi.....	9
1.6.2 Disseny.....	9
1.6.2 Conclusions.....	9
1.6.4 Línies d'evolució futura.....	9
2. ANÀLISI.....	10
2.1 Dimensions i Atributs.....	10
2.1.1 Anàlisi dels camps dels fitxers de dades.....	10
2.1.2 Anàlisi dels camps del fitxer “embalses.xls”.....	11
2.1.3 Anàlisi del quadre d'Estimacions de neu.....	11
2.2 Informes sol·licitats.....	11
2.3 Elements per a l'anàlisi multidimensional.....	11
2.3.1 Dimensions.....	11
2.3.2 Fets (Indicadors)	13
2.4 Diagrama del model conceptual	14
2.5 Diagrama de casos d'Ús	14
3. DISSENY.....	16
3.1 Disseny lògic.....	16
3.2 Disseny físic.....	16
3.3 Procés d'ETL.....	21
3.4 Recursos de Hardware i Software utilitzats.....	25
3.4.1 Hardware	25
3.4.2 Software.....	26
3.5 Disseny i descripció del informes creats.....	26
4. CONCLUSIONS.....	40
5. LÍNIES D'EVOLUCIÓ FUTURES.....	40
6. GLOSSARI.....	40
7. BIBLIOGRAFIA.....	41

1. Introducció

Aquest document és la memòria del Treball de Fi de Carrera dels estudis d'Enginyeria Tècnica d'Informàtica de Gestió. En aquest treball s'apliquen a un cas pràctic els coneixements teòrics acumulats durant aquests anys d'estudi.

La memòria explica amb detall les diferents fases del projecte realitzat per encàrrec de la Confederació Hidrogràfica Nord-Est.

1.1 Justificació del TFC i context en el qual es desenvolupa: punt de partida i aportació del TFC

La Confederació Hidrogràfica Nord-est (CHNE) necessita optimitzar i racionalitzar el seu procés d'obtenció d'informació. Per un altre banda també necessiten un conjunt d'informes que els ajudi en la presa de decisions.

Això s'aconseguirà automatitzant els processos de recollida de la informació i de generació d'informes. En aquest context neix la necessitat de la implementació d'un magatzem de dades corporatiu.

Amb aquest TFC es pretén donar resposta a les necessitat del client creant no només un magatzem de dades sinó una metodologia de treball .

Per un altre banda es tracta d'un producte que pot ser exportable a altres clients, com per exemple la resta de confederacions hidrogràfiques d'Espanya.

Per afrontar aquest projecte amb èxit cal partir d'un nivell mínim de coneixements en les següents àrees: Sistemes Gestors de Bases de Dades, gestió de projectes informàtics i Magatzem de Dades. També es necessari conèixer alguna eina de gestió de fulls de càlcul com MS Excel (Windows) o Calc de OpenOffice (Oracle)

Tot i partint d'una base, sempre és necessari aprendre les particularitats dels diferents sistemes. Això implica un procés d'aprenentatge amb la corresponent eina. En el meu cas jo coneixia Oracle Express , però no Oracle Discoverer i he hagut d'invertir un cert temps en conèixer l'eina.

1.2 Objectius del TFC

1.2.1 Vessant d'estudiant d'ETIG

Els principals objectius que es volen aconseguir mitjançant la realització d'aquest projecte són:

- Conèixer que són i com funcionen els Magatzems de Dades (DW).
- Conèixer el funcionament del programari **Oracle Express** (Sistema gestor de Bases de Dades) i d'**Oracle Discoverer** (Eina de Business Intelligence).
- Aprofundir en la creació de models de dades i d'informes.
- Tot i que no és un objectiu principal, el TFC també ens permetrà conèixer la virtualització de sistemes operatius gràcies a eines com **Virtualbox**.

Deixant de banda els aspectes purament tècnics amb aquest projecte també s'han d' assolir objectius de tipus formal i/o organitzatiu :

- Elaboració d'un conjunt de documents derivats del projecte: una memòria i una presentació.
- Aprendre a prioritzar en cas de conflicte entre tasques .
- Aprendre a planificar per assolir amb èxit les diferents fases del projecte.

1.2.2 Vessant Client

- Creació d'una eina flexible de permeti a la CHNE gestionar les seves dades
- Augmentar la eficiència del departament de planificació.
- Poder donar resposta ràpida a qualsevol petició d'informació.

1.3 Enfocament i Mètode Seguit

L'enfocament seguit ha estat totalment pràctic i basat en el cicle de desenvolupament clàssic:

- Anàlisi prèvia
- Definició del pla de treball
- Anàlisi funcional
- Disseny funcional
- Disseny tècnic
- Implementació
- Proves
- Lliurament

1.4 Planificació del Projecte

La planificació del Projecte es va dur a terme durant la PAC1 lliurada el dia 5 d'Octubre. En el document anomenat Pla de Treball es van definir els objectius a assolir, les tasques necessàries per poder aconseguir aquests objectius i la temporalitat d'aquestes tasques.

Donat que he hagut de refer aquest document, s'ha afegit el nou termini d'entrega per deixar constància d'aquest fet.

1.4.1 Tasques prèvies

Període: 20 al 22 de Setembre del 2010

- Descarregar i llegir L'enunciat del TFC de Magatzems de dades
- Descarregar i llegir el Pla Docent.
- Descarregar i instal·lar la màquina virtual. (aquesta tasca es podria fer posteriorment però és recomanable fer-ho quan abans millor per anticipar possibles problemes durant la instal·lació)

1.4.2 Realització PAC1

Període: 23 de Setembre al 5 d'Octubre del 2010

- Creació d'un pla de treball amb l'objectiu d'organitzar temporalment les tasques necessàries per realitzar amb èxit el Treball de Fi de Carrera.
- Creació d'un document preliminar on s'analitzen els requeriments i s'enumeren els principals elements d'anàlisi.

1.4.3 Realització PAC2

Període: 6 d'Octubre al 8 de Novembre del 2010

- Revisió i ampliació del document d'anàlisi dels requeriments entregat a la PAC 1.
- Disseny complet del producte tant a nivell conceptual com lògic
- Creació d'un esborrany del procés d'extracció, transformació i càrrega (ETL).
- Instal·lació i configuració del programari de gestió de bases de dades (Oracle).

Cal destacar que bona part de l'èxit del projecte dependrà d'aquesta fase. Un bon disseny ens estalviarà molt de temps en la fase d'implementació.

1.4.4 Realització PAC3

Període: 9 de Novembre al 20 de Desembre del 2010

- Creació de la base de dades i de les taules, vistes i usuaris necessaris.
- Càrrega de dades
- Instal·lació i configuració del programari d'explotació de dades (Oracle Discoverer)
- Creació i validació dels informes.
- Instal·lació de FileZilla, eina de transmissió de fitxers per FTP per tal de poder lliurar el producte.

Aquesta és la part del projecte en la que s'implementa el producte, es materialitzen els dissenys de la fase prèvia.

1.4.5 Lliurament final

Període primer lliurament: 21 de Desembre del 2010 al 10 de Gener del 2011

- Redacció de la memòria.
- Modificacions al producte.
- Preparació de la presentació audiovisual amb un programari de captura d'escriptori.

Cal destacar que la qualitat del document lliurat no va arribar als mínims exigits i se'm va recomanar la revisió i

ampliació d'aquest document.

Període segon lliurament: 29 de Maig del 2011 al 13 de Juny del 2011

Partint del document inicial s'ha procedit a la seva revisió, corregint les mancances ortogràfiques, d'expressió, de format i de contingut.

1.4.6 Debat Virtual final

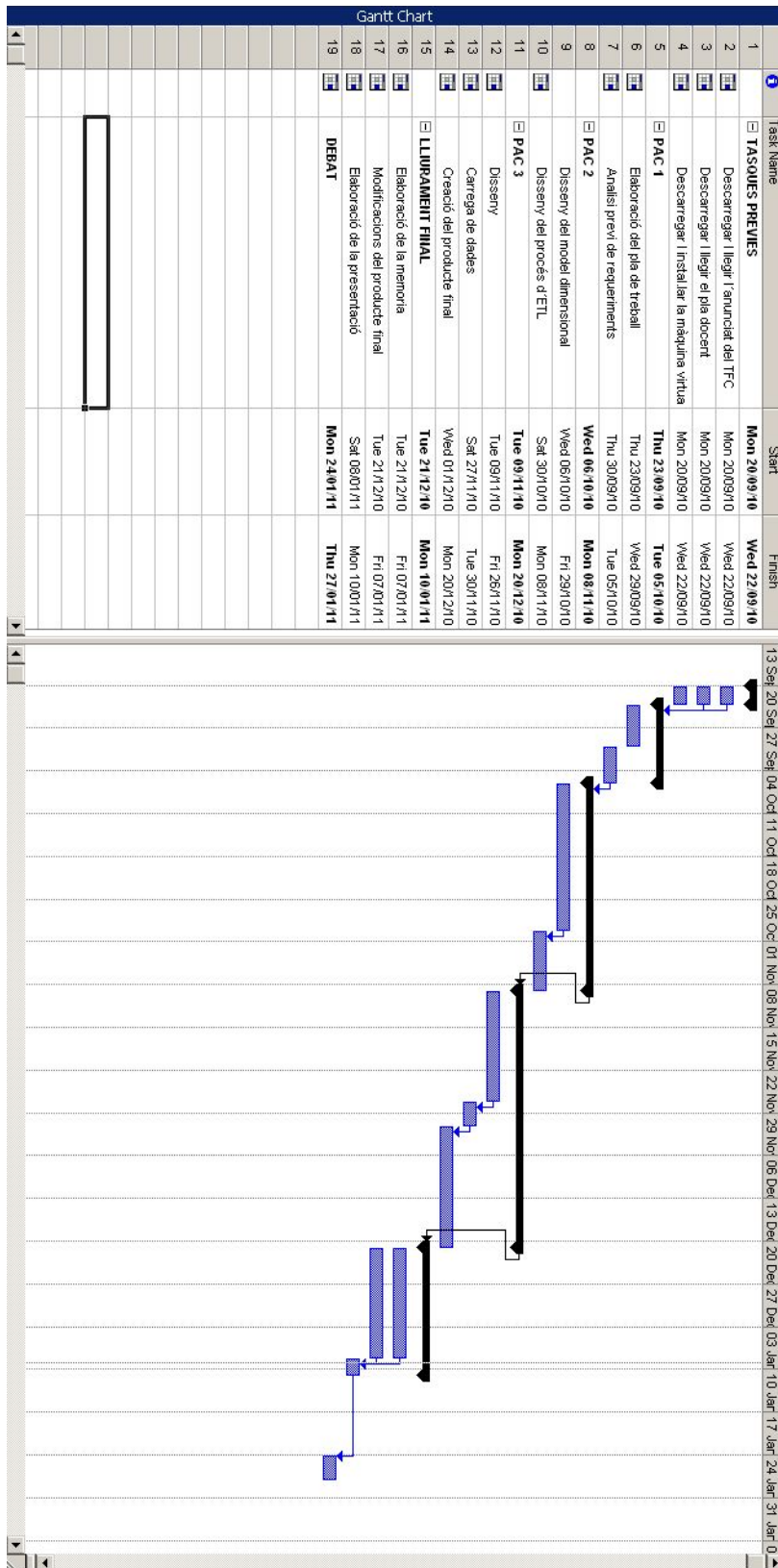
Període primer debat: 24 al 27 de Gener del 2011

Període segon debat: 20 al 23 de Juny del 2011

1.4.7 Diagrama de Gantt

S'adjunta a continuació un diagrama de Gantt que conté la informació següent:

- Tasques desenvolupades durant el projecte.
- Temps necessari per desenvolupar cadascuna d'aquestes tasques.
- Duració total del projecte.



Imatge 1. Diagrama de Gantt

1.5 Productes obtinguts

Durant el projecte s'han obtingut els següents productes:

Producte	Descripció
Pla de treball	Document lliurat a la PAC1. En el Pla de treball es defineixen les tasques necessàries per a realitzar el Treball de Fi de Carrera.
Anàlisi preliminar	Document lliurat a la PAC1. En aquest document es fa un primer anàlisi dels requeriments.
Anàlisi de requeriments	Document lliurat a la PAC2.
Instruccions per la càrrega de dades	Documentat lliurat a la PAC3 on es detallen els passos a seguir per realitzar la càrrega de dades.
Script en Unix	Aquest Script es necessita per generar els anomenats arxius de control (ctl) necessaris per la càrrega de dades.
Script en DOS	Aquest Script s'utilitza per fer la càrrega de dades de manera automàtica.
Arxius de Control	Arxius que necessita l'eina SQL Loader per fer la càrrega de dades.
Script SQL	Script en SQL de creació de les taules de la Base de Dades
Arxius .csv	Els arxius que contenen les dades sense transformar.
UOCPRACT.vdi	Es tracta del producte final. És un disc virtual per a Virtual Box que conté el magatzem de dades i els informes creats.
Memòria del Projecte	Document en format MS Word on es sintetitza el treball realitzat a les PACs.
Presentació audiovisual del Projecte	Realitzada en PowerPoint i gravada amb Camtasia Studio 7

1.6 Descripció dels altres capítols de la memòria

1.6.1 Anàlisi

Capítol destinat a l'anàlisi dels requeriments i al disseny funcional. S'expliquen les dimensions, atributs i indicadors principals.

1.6.2 Disseny

En aquest apartat s'aborda el disseny de la base de dades i de les arquitectures de maquinari i programari

1.6.3 Conclusions

En aquest apartat s'aborden les conclusions a les que s'arriba un cop finalitzat el projecte.

1.6.4 Línies d'evolució futura

Es comenten les possibles millores del producte final i també es contemplem les hipotètiques ampliacions d'aquest.

2. Anàlisi

2.1 Dades d'origen

La CHNE gestiona les seves dades en arxius de Microsoft Excel. Existeixen 54 fitxers que contenen les dades hidrogràfiques, i un fitxer anomenat "embalses.xls" que conté la localització geogràfica dels rius i embassaments. A més s'adjunta una taula que conté les estimacions de neu per embassament i any.

Tots els fitxers de dades tenen les següents característiques en comú:

- Les dades sempre es guarden en el full anomenat "Hoja1".
- El nom dels fitxers segueix sempre la següent convenció: *e+any+mes+dia*. Es tracta de la data en que s'analitza la situació dels embassaments, aquesta informació també es pot trobar al full " Hoja1".
- Contenen 10 camps que són els següents:

- 1) Embalses
- 2) Ríos
- 3) Capacidad Total Hm³
- 4) En esta fecha Hm³
- 5) % (1)
- 6) Entrada Hm³
- 7) Salida Hm³
- 8) Incremento Hm³
- 9) Hm³
- 10) % (2)

- Contenen 57 registres.

El fitxer "embalses.xls" conté els següents camps:

- 1) C.A.
- 2) Embalse
- 3) Río

2.1.1 Anàlisi dels camps dels fitxers de dades

- Embalses: Cadena de caràcters que conté el nom de l'embassament. Existeixen 57 embassaments diferents .
- Ríos: Cadena de caràcters que conté el nom del riu. Apareixen 42 rius diferents .
- Capacidad Total Hm³ : Valor numèric que conté la capacitat total del embassament expressada en Hm³.
- En esta fecha Hm³ : Valor numèric que conté la capacitat actual del embassament expressada en Hm³.
- % (1): Valor numèric calculat que conté la capacitat actual dividida de la capacitat total, el resultat de la divisió s'expressa com a percentatge.
- Entrada Hm³: Valor numèric que conté la aportació d'aigua al embassament expressada en Hm³.
- Salida Hm³: Valor numèric que conté la aportació d'aigua del embassament expressada en Hm³.
- Incremento Hm³: Valor numèric que conté l'augment o disminució d'aigua del embassament expressat en Hm³.
- Hm³ : Valor numèric que conté la capacitat del embassament durant el mateix període de l'any anterior expressada en Hm³.
- % (2): Valor numèric calculat que conté la capacitat de fa una any dividida de la capacitat total, el resultat de la divisió s'expressa com a percentatge.

2.1.2 Anàlisi dels camps del fitxer “embalses.xls”

- C.A.: Cadena de caràcters que conté una doble informació. Per una banda el nom de la província i per l'altre la comunitat autònoma a la que pertany aquesta província. Apareixen 6 Comunitats autònomes i 9 províncies.
- Embalse: Cadena de caràcters que conté el nom de l'embassament. Existeixen 41 embassaments diferents .
- Río: Cadena de caràcters que conté el nom del riu. Apareixen 31 rius diferents .

Després d'analitzar tots els fitxers s'han trobat inconsistències en la informació relativa als rius i embassaments. Hi ha embassaments que apareixen als fitxers de dades però no al fitxer d'embassaments i viceversa, existeixen noms que fan referència al mateix embassament però escrits de manera diferent i embassaments sense localització geogràfica.

Cal també destacar que els fitxers de dades contenen a vegades la cadena de caràcters '--' en comptes d'un valor nul.

Per solucionar el problema de la mancança d'informació geogràfica, s'ha fet servir informació provinent d'aquesta plana web <http://www.chebro.es/>.

La resta d'errades s'han corregit durant la fase d'extracció, transformació i càrrega (ETL).

2.1.3 Anàlisi del quadre d'Estimacions de neu

Aquest quadre mostra les estimacions per embassament i any. Les dades estan estructurades en forma de matriu, amb els anys, del 2001 al 2005, com a columnes i els embassaments com a files.

Cal destacar que no apareixen tots els embassaments considerats a la resta d'arxius. El quadre fa referència a 8 embassaments i agrupa la resta sota el títol RESTA.

2.2 Informes sol·licitats

LA CHNE ens demanen els següents informes:

- Capacitat total vs. capacitat actual en Hm3 per riu i any - mes
- Capacitat total vs. capacitat actual en Hm3 per embassament any - mes
- Capacitat total vs. capacitat actual en Hm3 per comunitat autònoma i any -mes
- Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments
- Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior (valors i % d'increment any anterior/any)
- Aportació (Entrada-Sortida) en Hm3 per riu, embassament i any -mes; comparativa amb la mitjana anual d'aportació de l'embassament
- Aigua diària en reserva disponible (Hm3/dia) a tota la confederació hidrogràfica per any – mes.

2.3 Elements per a l'anàlisi multidimensional

Un cop analitzades les dades i els informes a realitzar s'identifiquen els següents elements d'anàlisi:

2.3.1 Dimensions

Les dimensions representen els eixos sobre els que s'analitzarà la informació. S'han detectat les següents dimensions

- Dimensió Anys
- Dimensió Mes
- Dimensió Província

- Dimensió Embassament

Són necessàries 2 dimensions temporals autònomes. **Dimensió Mes** i **Dimensió Any**

DIMENSIÓ: ANYS	
ATRIBUTS	Descripció
Anys	Any d'anàlisi

DIMENSIÓ: MES	
ATRIBUTS	Descripció
Mes	Mes d'anàlisi
Nom_Mes	Nom del Mes

La dimensió que permetrà l'anàlisi geogràfica és la Província. Els atributs Província i Comunitat Autònoma presenten una relació jeràrquica, això implica que es poden agrupar les províncies en Comunitats Autònomes.

DIMENSIÓ: PROVINCIA	
ATRIBUTS	Descripció
ID	Identificador de la Província
PROVINCIA	Nom de la Província
COMUNITAT_AUTONOMA	Nom de la Comunitat Autònoma

Els atributs Embassament i Riu també tenen una relació jeràrquica, els diferents embassaments es poden agrupar en funció del riu al que pertanyen.

S'afegeix també l'atribut descriptor Província que facilitarà la càrrega i transformacions de dades.

DIMENSIÓ: EMBASSAMENT	
ATRIBUTS	Descripció
ID	Identificador de l'embassament
EMBASSAMENT	Nom de l'embassament

DIMENSIÓ: EMBASSAMENT	
RIU	Nom del Riu al que pertany l'embassament
PROVINCIA	Nom de la Província a la que pertany l'embassament

Així la partícula mínima d'informació, l'anomenat grànul , ha de contenir el següent grau de detall:

MES/ANY/PROVINCIA/EMBASSAMENT.

Nota explicativa:

Analitzant les dades s'observen les següents relacions:

-Donat 1 Embassament podem determinar a quina Província pertany.

-Donat 1 Embassament podem determinar a quin Riu pertany

-Donat 1 Riu aquest pot relacionar-se amb múltiples Províncies.

-Donat 1 Província aquesta és pot relacionar amb múltiples Rius.

Per tant el grànul podria quedar MES/ANY/ EMBASSAMENT i derivar RIU i PROVINCIA d'EMBASSAMENT .

Finalment s'ha optat per la primera opció, MES/ANY/PROVINCIA/EMBASSAMENT, tot i que PROVINCIA i EMBASSAMENT estan relacionades, corresponen a entitats diferents en el món real i tenen més sentit com a dimensions individuals.

2.3.2 Fets (Indicadors)

Els fets són l'aspecte quantificable o mesurable de les entitats que s'analitzen.

Un cop analitzades les dades s'han identificat els següents Fets:

- **TOTAL_HM3** : mesura la capacitat total de l'embassament.
- **ACTUAL_HM3** : mesura la capacitat actual de l'embassament.
- **ANTERIOR_HM3** : mesura la capacitat anterior de l'embassament.
- **ENTRADAS_HM3** : mesura les entrades d'aigua de l'embassament.
- **SALIDAS_HM3** : mesura les sortides d'aigua de l'embassament.
- **FC_MM_LTRS (*)**
- **NEU_EST (**)**

(*) FC_MM_Ltrs =Factor de Conversió a Milions de Litres.

Els indicadors sempre estaran expressats en Hm3 que serà la nostre unitat Standard de mesura. Un dels requeriments és poder veure la informació en Milions de Litres per això es mantindrà aquest factor de conversió a la nostra taula de fets.

FC_MM_Ltrs = 1000

doncs 1 Hm³ = Giga litre= 1.000.000.000 Litres

() Estimacio_Neu_HM3**

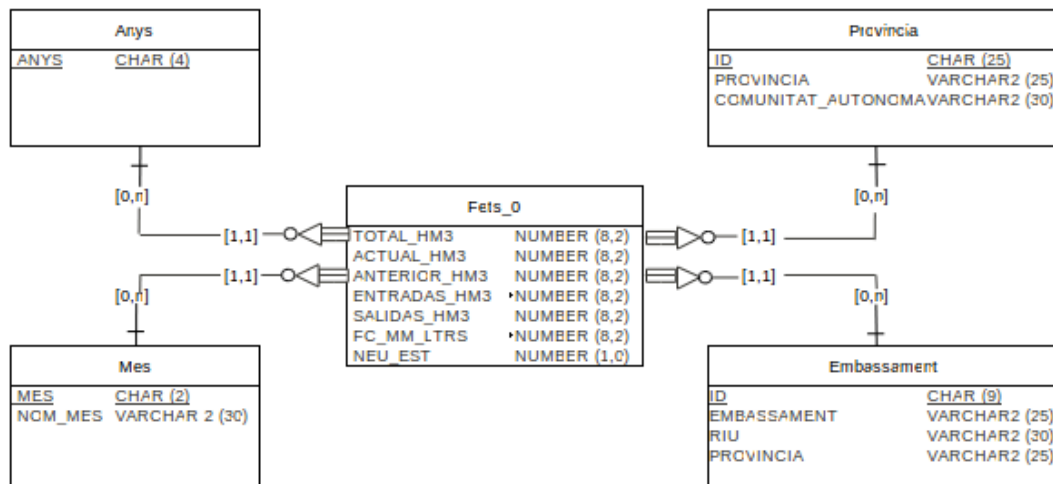
Es demana també informació relativa als **Hm3 d'aigua de neu equivalent** .Per obtenir aquesta magnitud seran necessàries les estimacions de neu en Hm3 que es mantindran a la taula de fets.

Nota explicativa 2:

Donat que la capacitat total és una informació fixe, es podria afegir com a atribut de l'embassament en comptes de com a indicador. Finalment per motius d'eficiència , es va decidir tractar aquesta informació com a indicador i no com un atribut.

2.4 Diagrama del model conceptual

S'inclou a continuació el model conceptual derivat de les dimensions i fets definits anteriorment. Es tracta del típic model en estrella, característic de molts projectes de Magatzem de dades.



Imatge 2. Diagrama del model conceptual

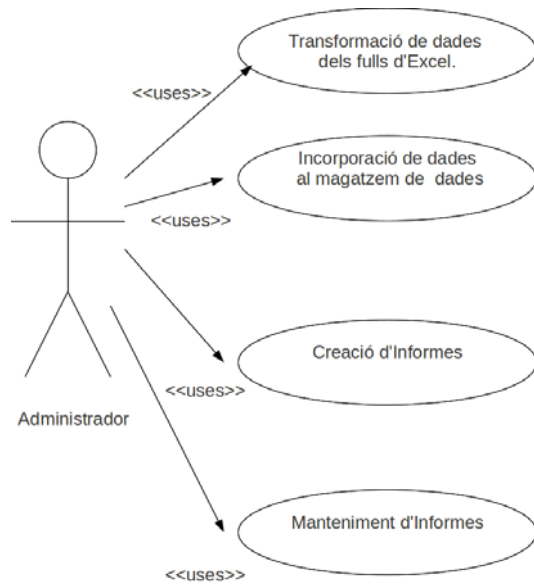
2.5 Diagrama de casos d'Ús

Per finalitzar aquest apartat es parlarà del diferents casos d'ús detectats.

S'han identificat 2 perfils d'usuari pel magatzem de dades de la CHNE:

1) Administrador.

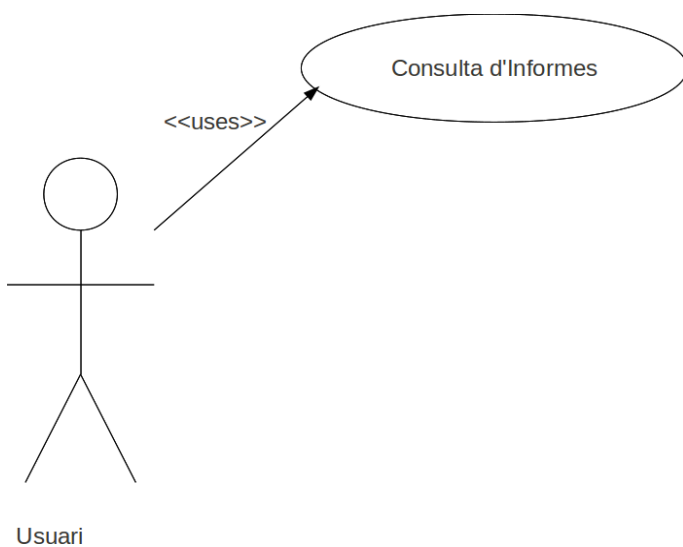
- Serà l'encarregat de fer la transformació de les dades provinents dels fulls d'Excel.
- Serà responsable de la incorporació de les dades transformades al magatzem de dades.
- Serà qui generi els informes .
- Serà qui s'encarregui del manteniment d'informes.



Imatge 4. Cas d'ús Administrador

2) L'usuari Final (tècnic usuari del data mart)

Serà l'encarregat d'executar i consultar els diferents informes.



Imatge 5. Cas d'ús Tècnic

3. Disseny

Partint del model conceptual creat anteriorment (veure punt 2.4) es procedeix a la creació dels models lògic i físic de la base de dades.

3.1 Disseny lògic

El disseny lògic s'obté creant una entitat per cadascuna de les dimensions definides i un altre per la taula de fets. Es necessari també identificar les claus d'unicitat, les claus foranes i les relacions entre les diferents entitats.

Entitats:

- **ANYS** (ANYS)
- **MES** (MES, NOM_MES)
- **PROVINCIA** (ID, PROVINCIA, COMUNITAT_AUTONOMA)
- **EMBASSAMENT** (ID, EMBASSAMENT, RIU, PROVINCIA)
- **FETS_0** (PROVINCIA_ID, EMBASSAMENT_ID, ANY_ID, MES_ID, TOTAL_HM3, ACTUAL_HM3, ANTERIOR_HM3, ENTRADAS_HM3, SALIDAS_HM3, FC_MM_LTRS, NEU_EST)

Relacions:

- ANYS.ANYS → FETS_0.ANY_ID
- MES.MES → FETS_0.MES_ID
- PROVINCIA.ID → FETS_0.PROVINCIA_ID
- EMBASSAMENT.ID → FETS_0.EMBASSAMENT_ID

3.2 Disseny físic

El model físic de la base de dades està format per les següents taules:

- **ANYS**: Conté els anys del 2001 al 2005, que són rellevants per aquest projecte. Aquesta taula s'actualitzarà anualment, donant d'alta un nou registre per l'any corresponent.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
ANYS	CHAR (4)	No	PK_ANY (Clau única)

- **MES**: Conté els 12 mesos de l'any, en format numèric (01, 02,...) i en format text (Gener, Febrer,...). En principi no està previst actualitzar aquesta taula. Si es volgués incorporar suport per diferents idiomes, s'afegirien noves columnes amb els noms dels mesos.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
MES	CHAR (2)	No	PK_MES (Clau única)
NOM_MES	VARCHAR2 (30)	Si	

- **PROVINCIA**: Conté les dades de les províncies necessàries per aquest projecte. La taula conté la relació existent entre província i comunitat autònoma.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
ID	CHAR (25)	No	PK_ID (Clau única)

PROVINCIA	VARCHAR2 (25)	No	NN_PROVINCIA (No Nul)
COMUNITAT_AUTONOMA	VARCHAR2 (30)	No	NN_COMUNITAT_AUTONOMA (No Nul)

Nota explicativa: Per motius de simplificació , s'ha optat per no crear una taula separada per les comunitats autònomes.

- **EMBASSAMENT:** Conté la informació relativa als embassaments necessaris per aquest projecte. Aquesta taula també conté la relació d'embassament amb riu i d'embassament amb província.

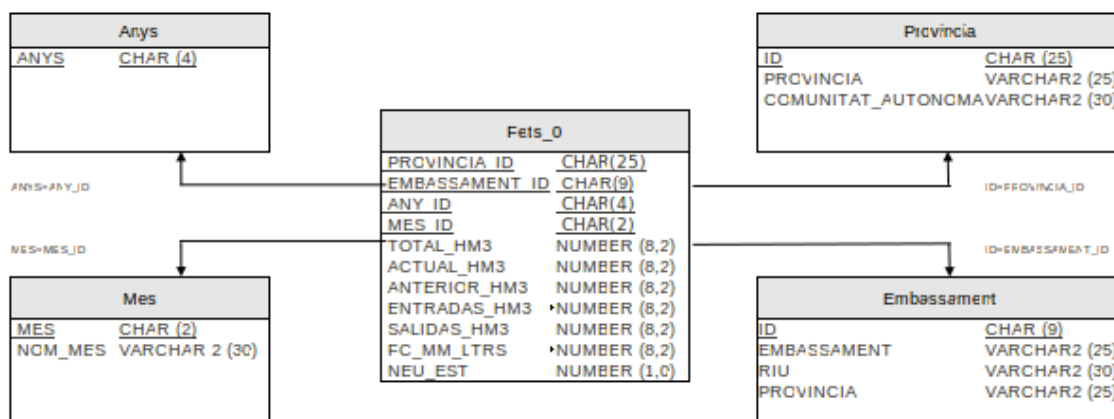
Nom de la Columna	Tipus de Dada	Nul	Restriccions
ID	CHAR (9)	No	PK_ANY (Clau única)
EMBASSAMENT	VARCHAR2 (25)	No	NN_EMBASSAMENT (No Nul)
RIU	VARCHAR2 (30)	No	NN_RIU (No Nul)
PROVINCIA	VARCHAR2 (25)	No	SYS_C004123 (No Nul).

Nota explicativa: Per motius de simplificació , s'ha optat per no crear una taula separada pels rius.

- **FETS_0:** Es tracta de la taula de fets. Aquesta taula conté tota la informació necessària per poder construir els informes. FETS_0 es relaciona amb la resta de les taules via clau forana.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
PROVINCIA_ID	CHAR (25)	No	PK_FETS_0 (Clau única)/ FK1_FETS_0 (Clau Forana)
EMBASSAMENT_ID	CHAR (9)	No	PK_FETS_0 (Clau única) / FK2_FETS_0 (Clau Forana)
ANY_ID	CHAR (4)	No	PK_FETS_0 (Clau única) / FK3_FETS_0 (Clau Forana)
MES_ID	CHAR (2)	No	PK_FETS_0 (Clau única) / FK4_FETS_0 (Clau Forana)
TOTAL_HM3	NUMBER (8,2)	Si	
ACTUAL_HM3	NUMBER (8,2)	Si	
ANTERIOR_HM3	NUMBER (8,2)	Si	
ENTRADAS_HM3	NUMBER (8,2)	Si	
SALIDAS_HM3	NUMBER (8,2)	Si	
FC_MM_LTRS	NUMBER (8,2)	Si	
NEU_EST	NUMBER (1,0)	Si	

Per últim s'adjunta un gràfic que explica de manera visual les diferents taules creades i les seves connexions.



Imatge 5. Model físic

Taules auxiliars

També ha estat necessari crear taules auxiliars o temporals per facilitar el procés de càrrega i transformació de les dades. Les taules creades són les següents:

- **TEMPORAL:** En aquesta taula es carreguen directament les dades provinents dels arxius d’Excel.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
CAMPO1	VARCHAR2 (25)	No	SYS_C004765 (Clau única)
CAMPO2	VARCHAR2 (25)	No	SYS_C004766 (Clau única)
CAMPO3	NUMBER (8,2)	Si	
CAMPO4	NUMBER (8,2)	Si	
CAMPO5	NUMBER (8,2)	Si	
CAMPO6	NUMBER (8,2)	Si	
CAMPO7	NUMBER (8,2)	Si	
CAMPO8	NUMBER (8,2)	Si	
CAMPO9	NUMBER (8,2)	Si	
CAMPO10	NUMBER (8,2)	Si	
CAMPO11	NUMBER (8,2)	Si	
CAMPO12	NUMBER (8,2)	Si	
CAMPO13	VARCHAR2 (25)	No	SYS_C004767 (Clau única)

- **NEU_TEMP**: En aquesta taula es carreguen directament les dades de l'arxiu d'estimacions de neu proporcionat per la CHNE.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
CAMPO1	VARCHAR2 (25)	Si	
CAMPO2	NUMBER (1,0)	Si	
CAMPO3	NUMBER (1,0)	Si	
CAMPO4	NUMBER (1,0)	Si	
CAMPO5	NUMBER (1,0)	Si	
CAMPO6	NUMBER (1,0)	Si	

- **NEU_DEF**: Aquesta taula és necessària per actualitzar la taula de fets FETS_0.

Nom de la Columna	Tipus de Dada	Nul	Restriccions
Embassament	VARCHAR2(25)	Si	
Neu_Est	NUMBER(1,0)	Si	
Any	VARCHAR2(25)	Si	

Scripts de creació de taules

El primer pas es crear un usuari a la base de dades. Aquest usuari crearà i carregarà les taules de la base de dades i definirà el model de dades.

Des de l'usuari system :

```
CREATE USER TEST_USER
IDENTIFIED BY rammstein
DEFAULT TABLESPACE users
QUOTA UNLIMITED ON users
TEMPORARY TABLESPACE temp;
```

```
GRANT CREATE SESSION,
CREATE TABLE
TO TEST_USER;
```

Es mostren a continuació els scripts utilitzats per crear les taules de la Base de Dades:

```
CREATE TABLE "MES"
(
  "MES" CHAR(2),
  "NOM_MES" VARCHAR2(30),
  CONSTRAINT "PK_MES" PRIMARY KEY ("MES") ENABLE
)
/
```

```
CREATE TABLE "PROVINCIA"
(
  "ID" CHAR(25),
```

```

"PROVINCIA" VARCHAR2(25 CHAR) CONSTRAINT "NN_PROVINCIA" NOT NULL ENABLE,
"COMUNITAT_AUTONOMA" VARCHAR2(30 CHAR) CONSTRAINT "NN_COMUNITAT_AUTONOMA" NOT NULL
ENABLE,
CONSTRAINT "PK_ID" PRIMARY KEY ("ID") ENABLE
)
/

```

```

CREATE TABLE "EMBASSAMENT"
(
  "ID" CHAR(9 CHAR),
  "EMBASSAMENT" VARCHAR2(25 CHAR) CONSTRAINT "NN_EMBASSAMENT" NOT NULL ENABLE,
  "RIU" VARCHAR2(30 CHAR) CONSTRAINT "NN_RIU" NOT NULL ENABLE,
  "PROVINCIA" VARCHAR2(25) NOT NULL ENABLE,
  CONSTRAINT "PK_ID2" PRIMARY KEY ("ID") ENABLE
)
/

```

```

CREATE TABLE "ANYS"
(
  "ANYS" CHAR(4 CHAR),
  CONSTRAINT "PK_ANY" PRIMARY KEY ("ANYS") ENABLE
)
/

```

```

CREATE TABLE FETS_0 (
Embassament_ID char (9 char) CONSTRAINT FK1_FETS_0 REFERENCES Embassament (ID),
Provincia_ID char (25 char) CONSTRAINT FK2_FETS_0 REFERENCES Provincia (ID),
Any_ID char (4 char) CONSTRAINT FK3_FETS_0 REFERENCES Anys (anys),
Mes_ID char (2 char) CONSTRAINT FK4_FETS_0 REFERENCES Mes (mes),
Total_HM3 NUMBER(8,2),
Actual_HM3 NUMBER(8,2),
Anterior_HM3 NUMBER(8,2),
Entradas_HM3 NUMBER(8,2),
Salidas_HM3 NUMBER(8,2),
FC_MM_Ltrs NUMBER(8,2),
Neu_Est NUMBER(1,0),
CONSTRAINT PK_FETS_0 PRIMARY KEY (Provincia_ID,Embassament_ID,Any_ID, Mes_ID );

```

```

CREATE TABLE "TEMPORAL"
(
  "CAMPO1" VARCHAR2(25) NOT NULL ENABLE,
  "CAMPO2" VARCHAR2(25) NOT NULL ENABLE,
  "CAMPO3" NUMBER(8,2),
  "CAMPO4" NUMBER(8,2),
  "CAMPO5" NUMBER(8,2),
  "CAMPO6" NUMBER(8,2),
  "CAMPO7" NUMBER(8,2),
  "CAMPO8" NUMBER(8,2),

```

```
"CAMPO9" NUMBER(8,2),
"CAMPO10" NUMBER(8,2),
"CAMPO11" NUMBER(8,2),
"CAMPO12" NUMBER(8,2),
"CAMPO13" VARCHAR2(25) NOT NULL ENABLE
)
/
```

```
CREATE TABLE "NEU_TEMP"
(
  "CAMPO1" VARCHAR2(25),
  "CAMPO2" NUMBER(1,0),
  "CAMPO3" NUMBER(1,0),
  "CAMPO4" NUMBER(1,0),
  "CAMPO5" NUMBER(1,0),
  "CAMPO6" NUMBER(1,0)
)
/
```

```
CREATE TABLE "NEU_DEF"
(
  "Embassament" VARCHAR2(25),
  "Neu_Est" NUMBER(1,0),
  "Any" VARCHAR2(25)
)
/
```

3.3 Procés d'ETL

L'eina ETL que es fa servir per la càrrega de les dades és el SQL Loader. A continuació es descriuen els passos necessaris per realitzar la càrrega de dades.

Es necessari transformar els arxius de format Excel a format .csv (valors separats per comes), que és el format requerit per l'eina SQL Loader.

Aquesta transformació s'ha fet obrint cadascun dels arxius de dades amb l'eina Calc i guardant els arxius amb la nova extensió .csv.

Adicionalment es creen els següents fitxers .csv :

- Provincia.csv: aquest arxiu es genera transformant l'arxiu "embalses.xls" . S'ha fet servir la informació de la columna A referent a província i comunitat autònoma. També s'ha afegit una nova columna amb el codi ISO de la Província.
- Embassament.csv: aquest arxiu es genera transformant qualsevol dels arxius de dades. S'han fet servir les columnes A i B que fan referència a embassament i riu.
L'arxiu s'ha complementat amb informació provinent de "embalses.xls" i del enllaç mencionat al punt 2.1.2 d'aquest document. Per últim s'ha afegit una columna amb la informació de província.
- Anys.csv: arxiu que conté la informació referent als anys, s'ha generat manualment degut a la seva baixa complexitat.
- Mes.csv: arxiu que conté la informació referent als mesos, s'ha generat manualment degut a la seva baixa complexitat.
- Neu.csv: aquest arxiu s'ha creat fent servir el quadre d'estimacions de neu proporcionat per la CHNE. No han calgut transformacions prèvies.

El procés de càrrega de les taules Provincia, Embassament, Anys, Mes i Neu es idèntic. Per carregar la taula de fets FETS_0 serà necessari un procés més complex, que s'explicarà més endavant.

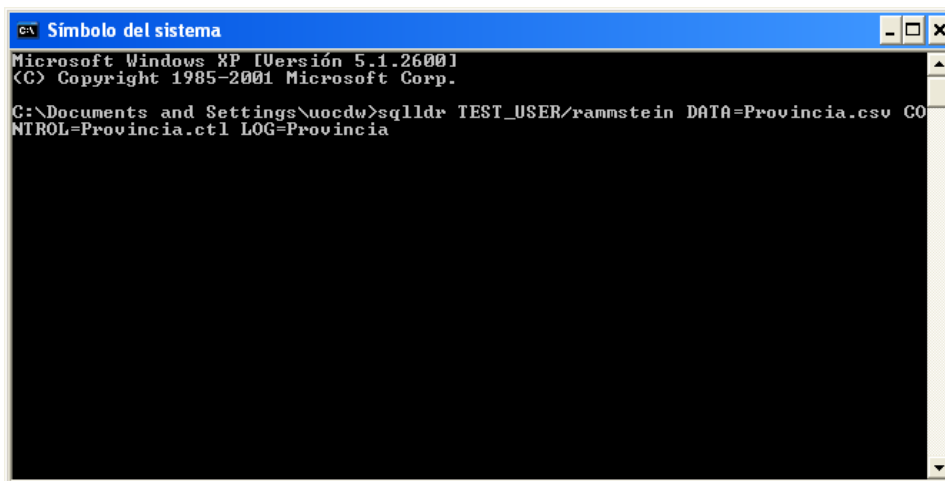
Per explicar el primer procés de càrrega es fa servir la taula provincia com a exemple.

El primer pas es crear l'arxiu de control, aquest arxiu és un tipus especial d'arxiu, que fa servir l'eina SQL Loader per procesar les dades.

L'arxiu de control provincia.ctl:

```
load data
CHARACTERSET UTF8
infile 'provincia.csv'
into table provincia fields terminated by "," optionally enclosed by ""
( ID, Provincia, Comunitat_Autonoma )
```

A continuació s'executa el procés de càrrega des de la consola de DOS. Es necessària la següent informació, nom i password de l'usuari de la Base de Dades, noms dels arxius de dades i de control



```
Símbolo del sistema
Microsoft Windows XP [Versión 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\uocdu>sqlldr TEST_USER/rammstein DATA=Provincia.csv CONTROL=Provincia.ctl LOG=Provincia
```

Imatge 6. SQL Loader

L'arxiu provincia.log ajuda a monitoritzar el procés de càrrega. La imatge següent mostra com els 11 registres han estat carregats amb èxit.

```

provincia.log - Bloc de notas
Archivo Edición Formato Ver Ayuda

SQL*Loader: Release 10.2.0.1.0 - Production on Sáb Dic 11 19:02:38 2010
Copyright (c) 1982, 2005, oracle. All rights reserved.

Archivo de Control:  provincia.ct1
Juego de caracteres UTF8 especificado para todas las entradas.
El primer archivo de datos primario Provincia.csv tiene una
marca de orden de bytes utf8.

Archivo de Datos:      Provincia.csv
Archivo de Errores:   Provincia.bad
Desechar Archivo:    ninguno especificado

(Permitir todos los registros desechados)

Número a cargar: ALL
Número a ignorar: 0
Errores permitidos: 50
Matriz de enlace:    64 filas, máximo de 256000 bytes
Continuación:        ninguno especificado
Ruta de acceso utilizada:  convencional

Tabla PROVINCIA, cargada de cada registro lógico.
Opción INSERT activa para esta tabla: INSERT

-----
Nombre Columna                Posición  Long  Term Entorno Tipo de Dato
-----
ID                             FIRST     *    ,  O('') CHARACTER
PROVINCIA                      NEXT     *    ,  O('') CHARACTER
COMUNITAT_AUTONOMA            NEXT     *    ,  O('') CHARACTER
-----

Tabla PROVINCIA:
 11 Filas se ha cargado correctamente.
 0 Filas no cargada debido a errores de datos.
 0 Filas no cargada porque todas las cláusulas WHEN han fallado.
 0 Filas no cargada porque todos los campos eran nulos.

Espacio asignado a matriz de enlace:          49536 bytes (64 filas)
Bytes de buffer de lectura: 1048576

Total de registros lógicos ignorados:         0
Total de registros lógicos leídos:           11
Total de registros lógicos rechazados:        0
Total de registros lógicos desechados:        0

La ejecución empezó en Sáb Dic 11 19:02:38 2010
La ejecución terminó en Sáb Dic 11 19:02:38 2010

```

Imatge 7. Fitxer log.

Es repeteixen aquests passos per la resta de taules a excepció de la taula de fets.

Per poder processar els 54 fitxers de dades i actualitzar la taula de fets FETS_0 el procediment serà diferent. Les dades es carreguen inicialment a la taula auxiliar TEMPORAL.

El primer pas es crear un arxiu genèric de control.

```

OPTIONS (SKIP=9)
load data
CHARACTERSET UTF8
infile ":FILE"
Append into table TEMPORAL when Campo1 <> "" fields terminated by "," optionally enclosed by "" trailing nullcols

(
Campo1 ,
Campo2 ,
Campo3 ,
Campo4 ,
Campo5 ,
Campo6 ,
Campo7 "replace(:Campo7,'---',0)",
Campo8 ,

```



```
Campo9 ,
Campo10,
Campo11,
Campo12,
Campo13 CONSTANT ":FILE"
)
```

L'objectiu es d'una banda poder fer servir la informació d'any i mes continguda al nom del arxius .csv i per l'altre generar tants arxius de control com arxius de dades (.csv) hi hagi.

Per aconseguir-ho s'ha fet servir el següent Shell Script:

```
FILES=`ls *.csv`
CTL=GENERIC.ctl
for f in $FILES
do
  cat $CTL | sed "s/:FILE/$f/g" > $f.ctl
done
```

Per últim es crea un fitxer anomenat script .bat que carregarà a la taula TEMPORAL els 54 fitxers de dades.

Durant la càrrega es realitzen una sèrie de transformacions per "netejar" al màxim les dades. Aquestes transformacions es fan mitjançant instruccions dels fitxers de control.

- Amb la instrucció *skip = 9* s'ignoren les 9 primeres línies de cada fitxer, aquestes línies no contenen informació rellevant.
- Amb la instrucció *when Campo1 <> ""* es filtren les últimes línies de cada fitxer, aquestes línies contenen informació que no és necessària.
- Amb la instrucció *"replace(:Campo7,'---',0)"* es netegen els arxius de caràcters espuris ("---") substituint aquests per zeros.

Abans de poder actualitzar la taula FETS_0 amb la informació carregada a la taula TEMPORAL caldrà fer un últim pas, actualitzar la taula NEU_DEF.

La taula NEU_TEMP no té l'estructura necessària per poder actualitzar la taula de fets FETS_0. La següent consulta inserta les dades de NEU_TEMP a NEU_DEF, que si té l'estructura desitjada.

```
insert into NEU_DEF
select UPPER(CAMPO1) as Embassament ,CAMPO2 as Neu, 2001 as "Any" from NEU_TEMP
UNION
select UPPER(CAMPO1) as Embassament ,CAMPO3 as Neu, 2002 as "Any" from NEU_TEMP
UNION
select UPPER(CAMPO1) as Embassament ,CAMPO4 as Neu, 2003 as "Any" from NEU_TEMP
UNION
select UPPER(CAMPO1) as Embassament ,CAMPO5 as Neu, 2004 as "Any" from NEU_TEMP
UNION
select UPPER(CAMPO1) as Embassament ,CAMPO6 as Neu, 2005 as "Any" from NEU_TEMP
;
```

Un cop actualitzada la taula NEU_DEF, es pot actualitzar finalment la taula de FETS_0 amb la següent consulta:

Insert into FETS_0

```
select Embassament.ID AS "EMBASSAMENT_ID",PROVINCIA.ID AS "PROVINCIA_ID", substr (CAMPO13,2,4) as "ANY"
,substr (CAMPO13, 6,2) as "MES",CAMPO3 as "Total_HM3", Campo4 as "Actual_HM3",CAMPO11 as
"Anterior_HM3",CAMPO6 as "Entradas_HM3",CAMPO7 as "Salidas_HM3", 1000 as
"FC_MM_Ltrs",NVL(NEU_DEF."Neu_Est",1) AS "Neu_Est"
```

```
from ((TEMPORAL INNER JOIN EMBASSAMENT ON TEMPORAL.CAMPO1=EMBASSAMENT.EMBASSAMENT) INNER JOIN
PROVINCIA ON EMBASSAMENT.PROVINCIA=PROVINCIA.PROVINCIA)
LEFT OUTER JOIN NEU_DEF ON (TEMPORAL.CAMPO1=NEU_DEF."Embassament" AND substr
(TEMPORAL.CAMPO13,2,4)=NEU_DEF."Any" )
;
```

Amb aquesta consulta s'han fet les següents transformacions:

- Mitjançant la funció SQL **substr** s'han derivat l'any i el mes del camp "CAMPO13"
- S'ha substituït el valor NULL per un 1 en el camp Neu_Est, en aquells registres on el camp embassament té el valor "RESTA". Per aconseguir-ho s'ha fet servir la funció NVL

Es important destacar que no s'han proporcionat fitxers de dades pel període de Gener a Juny del 2001. Es pot recuperar la informació relativa a les capacitats actuals gràcies al camp ANTERIOR de les dades de Gener a Juny del 2002.

Mitjançant la següent consulta s'afegiran aquestes dades a la taula de fets.

INSERT INTO FETS_0

```
select EMBASSAMENT_ID, PROVINCIA_ID, 2001,MES_ID,
TOTAL_HM3,ANTERIOR_HM3,0,0,0,FC_MM_LTRS,NVL(NEU_DEF."Neu_Est",1) AS "Neu_Est"
```

```
from (FETS_0 INNER JOIN EMBASSAMENT ON FETS_0.EMBASSAMENT_ID=EMBASSAMENT.ID)
```

```
LEFT OUTER JOIN NEU_DEF ON (EMBASSAMENT.EMBASSAMENT=NEU_DEF."Embassament" AND
NEU_DEF."Any"='2001')
```

```
where ANY_ID='2002' and (MES_ID='01' or MES_ID='02' or MES_ID='03' or MES_ID='04' or MES_ID='05' or
MES_ID='06');
```

3.4 Recursos de Hardware i Software utilitzats

3.4.1 Hardware

S'han fet servir 2 ordinadors portàtils connectats a una mateixa xarxa wi-fi:

- Un ASUS PRO 60 amb processador Intel Core 2 a 1.66 GHz i 1 GB de memòria RAM , que s'ha fet servir per realitzar les tasques directament relacionades amb el projecte:
 - Instal·lació de la màquina virtual amb els programaris de Base de Dades i de Business Intelligence.
 - Transformació i càrrega de les dades.
 - Creació d'informes.

- Un MacBook Air amb processador Intel Core 2 Duo a 1.86 GHz i 2 GB de memòria RAM a 1067 MHz , que s'ha fet servir com a ordinador secundari per realitzar les tasques de documentació del projecte:
 - Creació de la memòria del projecte.
 - Creació de la presentació audiovisual.

3.4.2 Software

Sistemes Operatius

L'ordinador principal està equipat amb Ubuntu 10.4 Lucid Lynx i el secundari amb Mac OS X Snow Leopard. Malgrat això el sistema operatiu que s'ha fet servir ha estat el Windows XP, doncs era l'instal·lat a la màquina virtual proporcionada pel client.

A l'ordinador secundari també s'ha fet servir una màquina virtual amb Windows XP per poder utilitzar tant MS Word com MS PowerPoint necessaris per generar la documentació

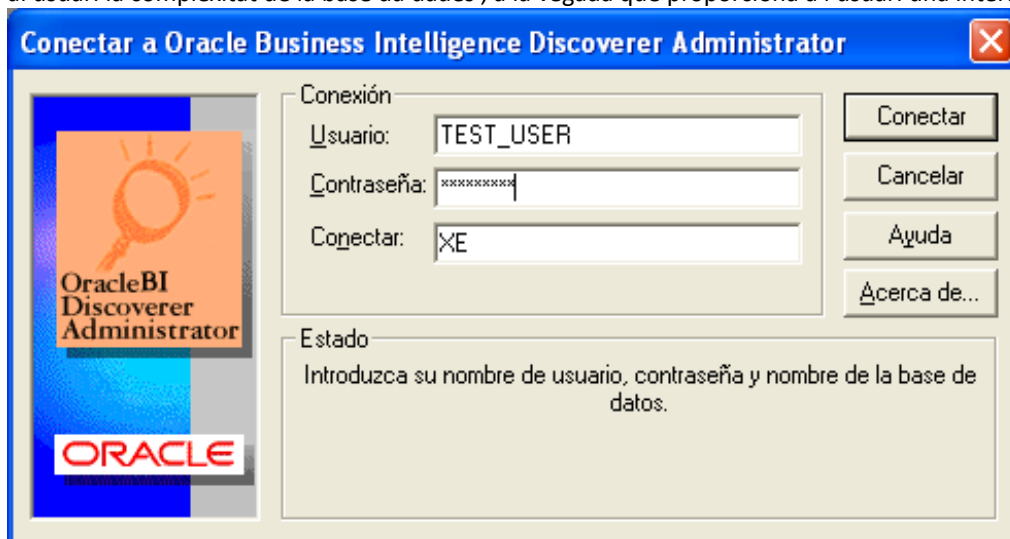
Programari

- Oracle Database 10 g Express Edition : és el programari de gestió de base de dades.
- Oracle Discoverer Administrator: eina per gestionar l'univers de dades i construir els informes.
- Oracle Discoverer Desktop: eina per visualitzar i fer el manteniment d'informes.
- MS Office 2010 de Windows: s'ha fet servir per generar la documentació en MS Word i MS PowerPoint. També s'ha fet servir MS Project per crear el diagrama de Gantt.
- Open Office d'Oracle : s'ha fet servir per gestionar els fulls de càlcul amb les dades originals.
- Dropbox: per compartir arxius de forma ràpida i fer còpies de seguretat dels arxius més importants.
- Camtasia Studio 7 y Win FF: s'ha fet enregistrar la presentació audiovisual i convertir-la a format .flv

3.5 Disseny i descripció del informes creats

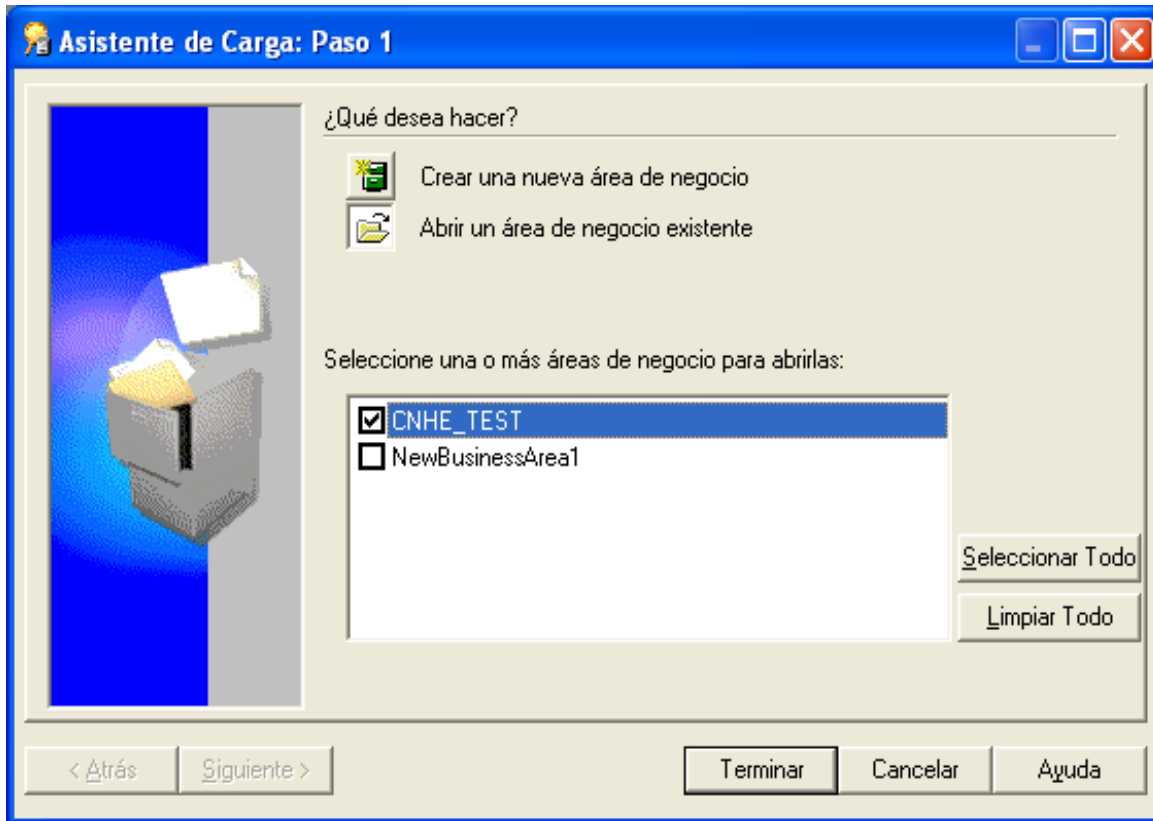
Per dissenyar i consultar els informes s'ha fet servir l'eina de Business Intelligence Oracle Discoverer.

Amb aquesta eina s'ha creat l'anomenat EUL (END USER LAYER) o capa final d'usuari, es tracta d'una capa per amagar al usuari la complexitat de la base da dades , a la vegada que proporciona a l'usuari una interfície fàcil d'entendre.

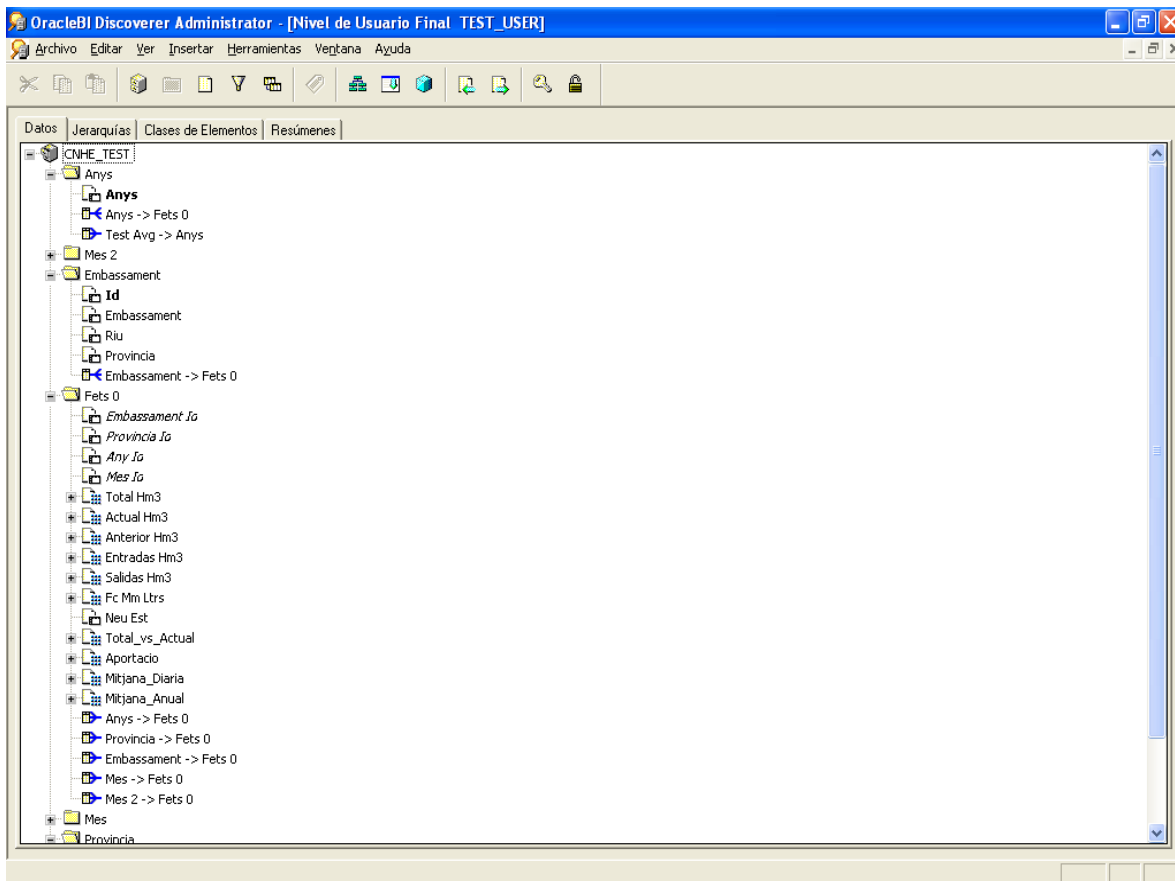


Imatge 8. Connexió Oracle Business Intelligence Discoverer Administrator

A continuació s'ha creat l'àrea de negoci CHNE_TEST. L'àrea de negoci és una agrupació conceptual de taules i vistes que s'adapten als requeriments específics d'informació de l'usuari final. Els informes creats estaran vinculats a aquesta àrea de negoci.



Imatge 9. Àrea de negoci



Imatge 10. EUL

Des de la base de dades s'ha creat un nou usuari, TEST_USER2, que podrà crear informes fent servir el Oracle Business Intelligence Discoverer Desktop.

Des de l'usuari system s'ha fet :

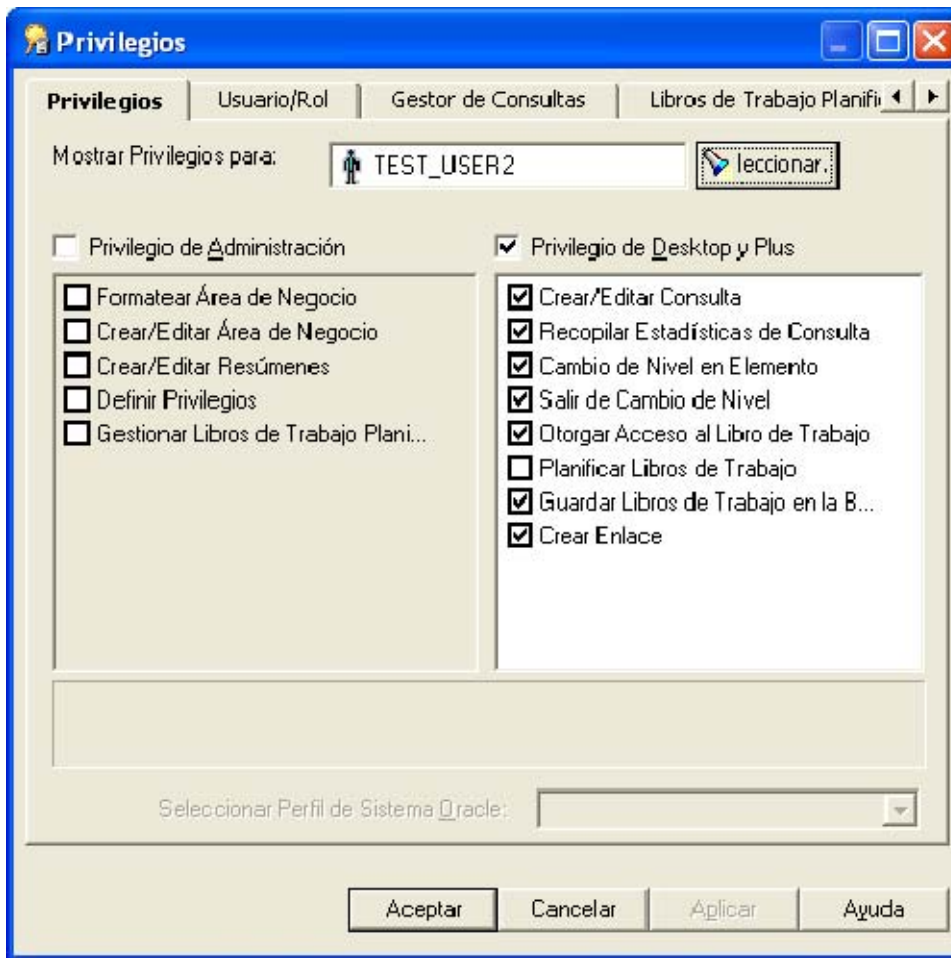
```
CREATE USER TEST_USER2
IDENTIFIED BY metallica
DEFAULT TABLESPACE users
QUOTA UNLIMITED ON users
TEMPORARY TABLESPACE temp;
```

```
GRANT CREATE SESSION
TO TEST_USER2;
```

Desde l'usuari TEST_USER s'ha donat accés de lectura a les taules necessàries :

```
GRANT SELECT ON NOM_TAUOLA TO TEST_USER2 WITH GRANT OPTION;
```

i desde el Discoverer Administrator se li han donat drets per crear i editar informes:

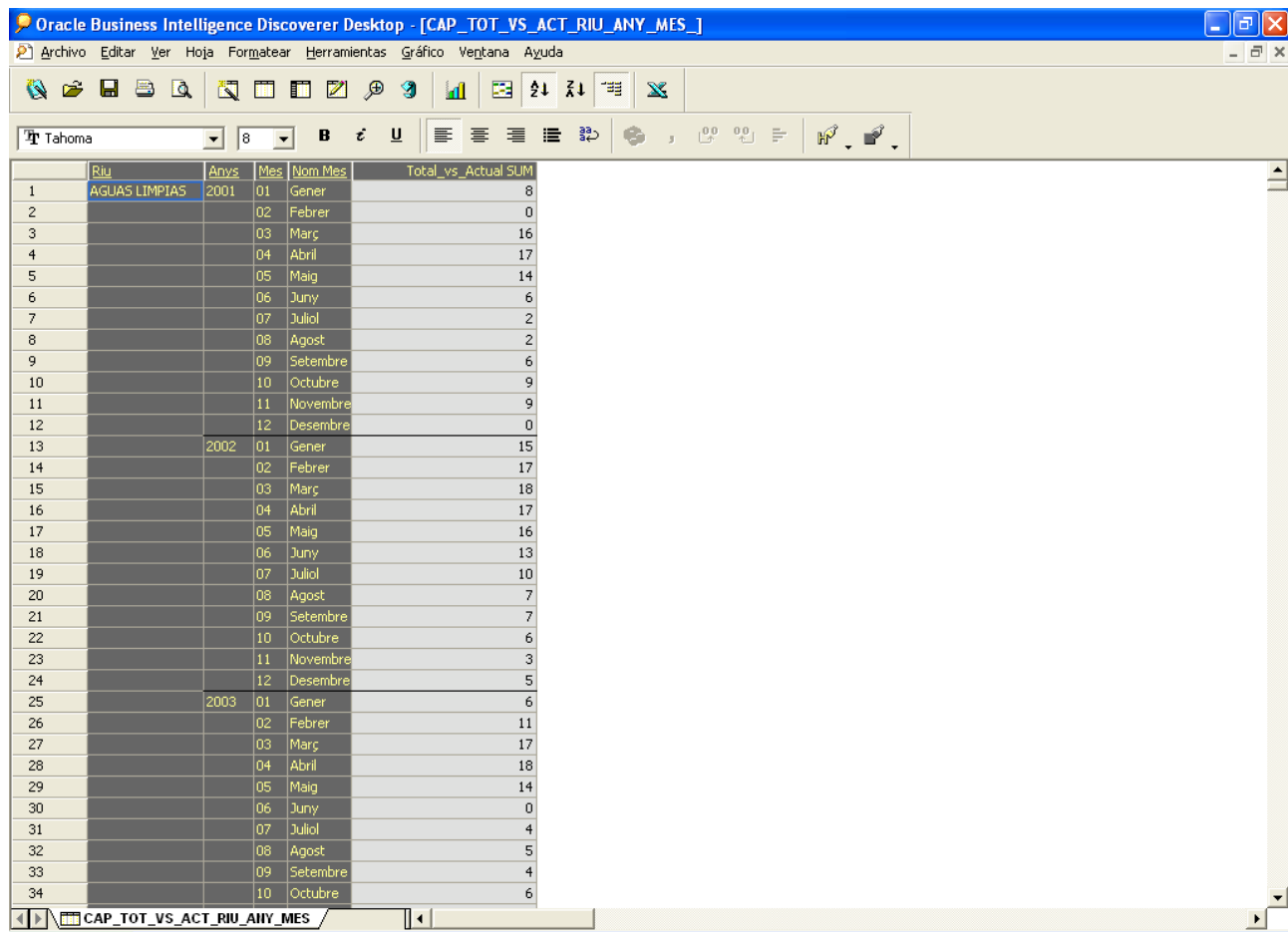


Imatge 11. Permisos usuari TEST_USER2.

Aquests són els principals informes demanats pel client:

Necessitats d'informació	Informes Proposats
<ul style="list-style-type: none"> • Capacitat total vs. capacitat actual en Hm3 per riu i any - mes 	<u>CAP TOT VS ACT RIU ANY MES</u>
<ul style="list-style-type: none"> • Capacitat total vs. capacitat actual en Hm3 per embassament any - mes 	<u>CAP TOT VS ACT EMB ANY MES</u>
<ul style="list-style-type: none"> • Capacitat total vs. capacitat actual en Hm3 per comunitat autònoma i any - mes 	<u>CAP TOT VS ACT CCAA ANY MES</u>
<ul style="list-style-type: none"> • Comparativa anual de l'evolució per comunitat autònoma de la capacitat agregada de tots els seus embassaments 	<u>COMP ANUAL EVOL CCAA</u>
<ul style="list-style-type: none"> • Comparativa de la capacitat mitjana agregada de tots els embassaments d'una comunitat autònoma entre un any i l'any anterior (valors i % d'increment any anterior/any) 	<u>COMP CAP MITJ CCAA INTERANUAL</u>
<ul style="list-style-type: none"> • Aportació (Entrada-Sortida) en Hm3 per riu, embassament i any - mes; comparativa amb la mitjana anual d'aportació de l'embassament 	<u>APORT RIU EMB ANY MES CMP MITJ</u>
<ul style="list-style-type: none"> • Aigua diària en reserva disponible (Hm3/dia) a tota la confederació hidrogràfica per any - mes 	<u>AIGUA DIA DISP ANY MES</u>

CAP TOT VS ACT RIU ANY MES

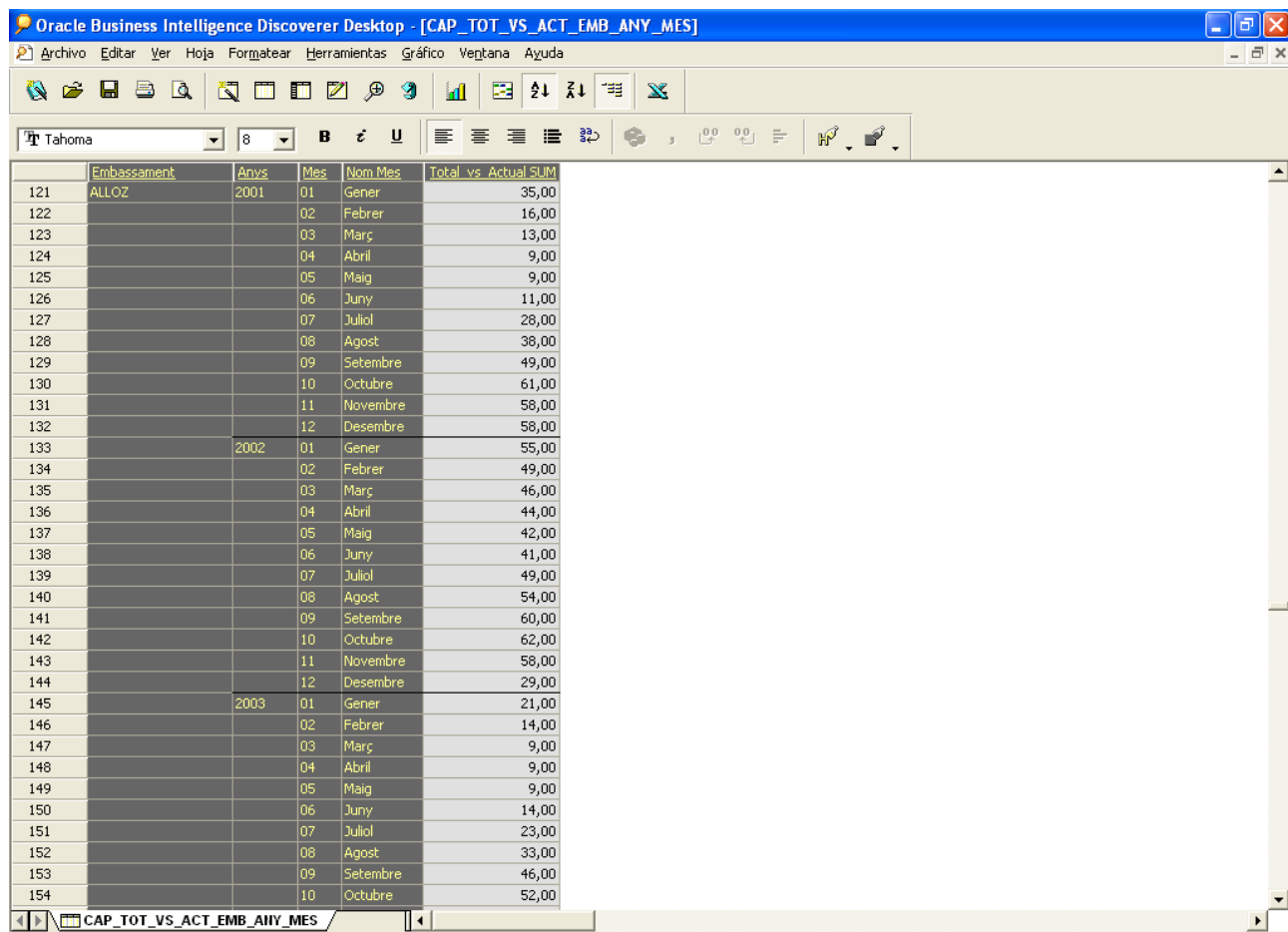


The screenshot shows the Oracle Business Intelligence Discoverer Desktop interface. The main window displays a table report with the following data:

	Riu	Anys	Mes	Nom Mes	Total_vs_Actual SUM
1	AGUAS LIMPIAS	2001	01	Gener	8
2			02	Febrer	0
3			03	Març	16
4			04	Abril	17
5			05	Maig	14
6			06	Juny	6
7			07	Juliol	2
8			08	Agost	2
9			09	Setembre	6
10			10	Octubre	9
11			11	Novembre	9
12			12	Desembre	0
13		2002	01	Gener	15
14			02	Febrer	17
15			03	Març	18
16			04	Abril	17
17			05	Maig	16
18			06	Juny	13
19			07	Juliol	10
20			08	Agost	7
21			09	Setembre	7
22			10	Octubre	6
23			11	Novembre	3
24			12	Desembre	5
25		2003	01	Gener	6
26			02	Febrer	11
27			03	Març	17
28			04	Abril	18
29			05	Maig	14
30			06	Juny	0
31			07	Juliol	4
32			08	Agost	5
33			09	Setembre	4
34			10	Octubre	6

Imatge 12. Informe 1

CAP TOT VS ACT EMB ANY MES

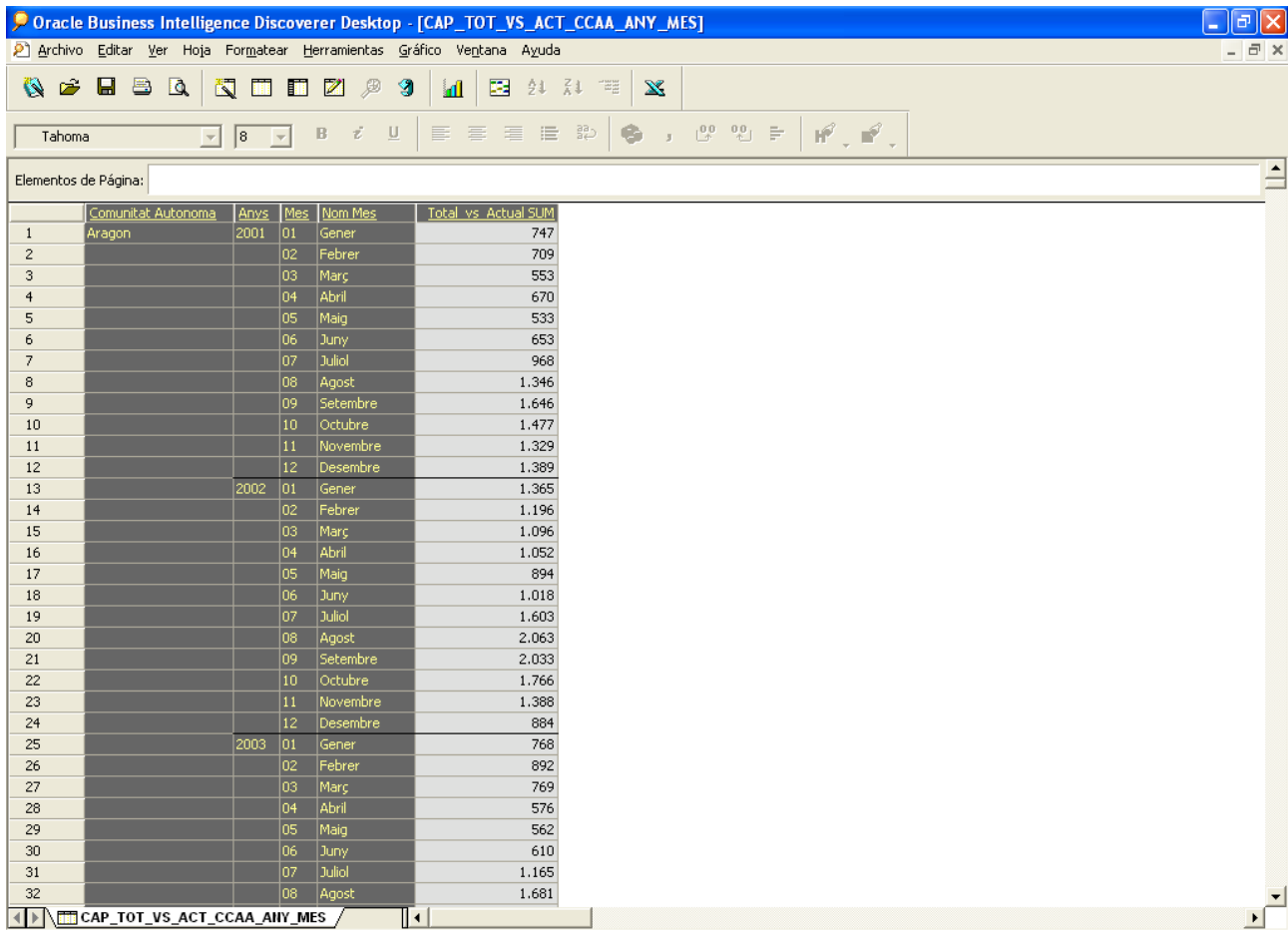


The screenshot shows the Oracle Business Intelligence Discoverer Desktop interface. The main window displays a data table with the following columns: 'Embassament', 'Anys', 'Mes', 'Nom Mes', and 'Total vs. Actual SUM'. The data is organized by year (2001, 2002, 2003) and then by month (01 to 12). The first row (121) shows a value of 35,00 for the month of January (Gener) in 2001. The table continues down to row 154, which shows a value of 52,00 for the month of October (Octubre) in 2003.

	Embassament	Anys	Mes	Nom Mes	Total vs. Actual SUM
121	ALLOZ	2001	01	Gener	35,00
122			02	Febrer	16,00
123			03	Març	13,00
124			04	Abril	9,00
125			05	Maig	9,00
126			06	Juny	11,00
127			07	Juliol	28,00
128			08	Agost	38,00
129			09	Setembre	49,00
130			10	Octubre	61,00
131			11	Novembre	58,00
132			12	Desembre	58,00
133		2002	01	Gener	55,00
134			02	Febrer	49,00
135			03	Març	46,00
136			04	Abril	44,00
137			05	Maig	42,00
138			06	Juny	41,00
139			07	Juliol	49,00
140			08	Agost	54,00
141			09	Setembre	60,00
142			10	Octubre	62,00
143			11	Novembre	58,00
144			12	Desembre	29,00
145		2003	01	Gener	21,00
146			02	Febrer	14,00
147			03	Març	9,00
148			04	Abril	9,00
149			05	Maig	9,00
150			06	Juny	14,00
151			07	Juliol	23,00
152			08	Agost	33,00
153			09	Setembre	46,00
154			10	Octubre	52,00

Imatge 13. Informe 2

CAP TOT VS ACT CCAA ANY MES



Oracle Business Intelligence Discoverer Desktop - [CAP_TOT_VS_ACT_CCAA_ANY_MES]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B z U

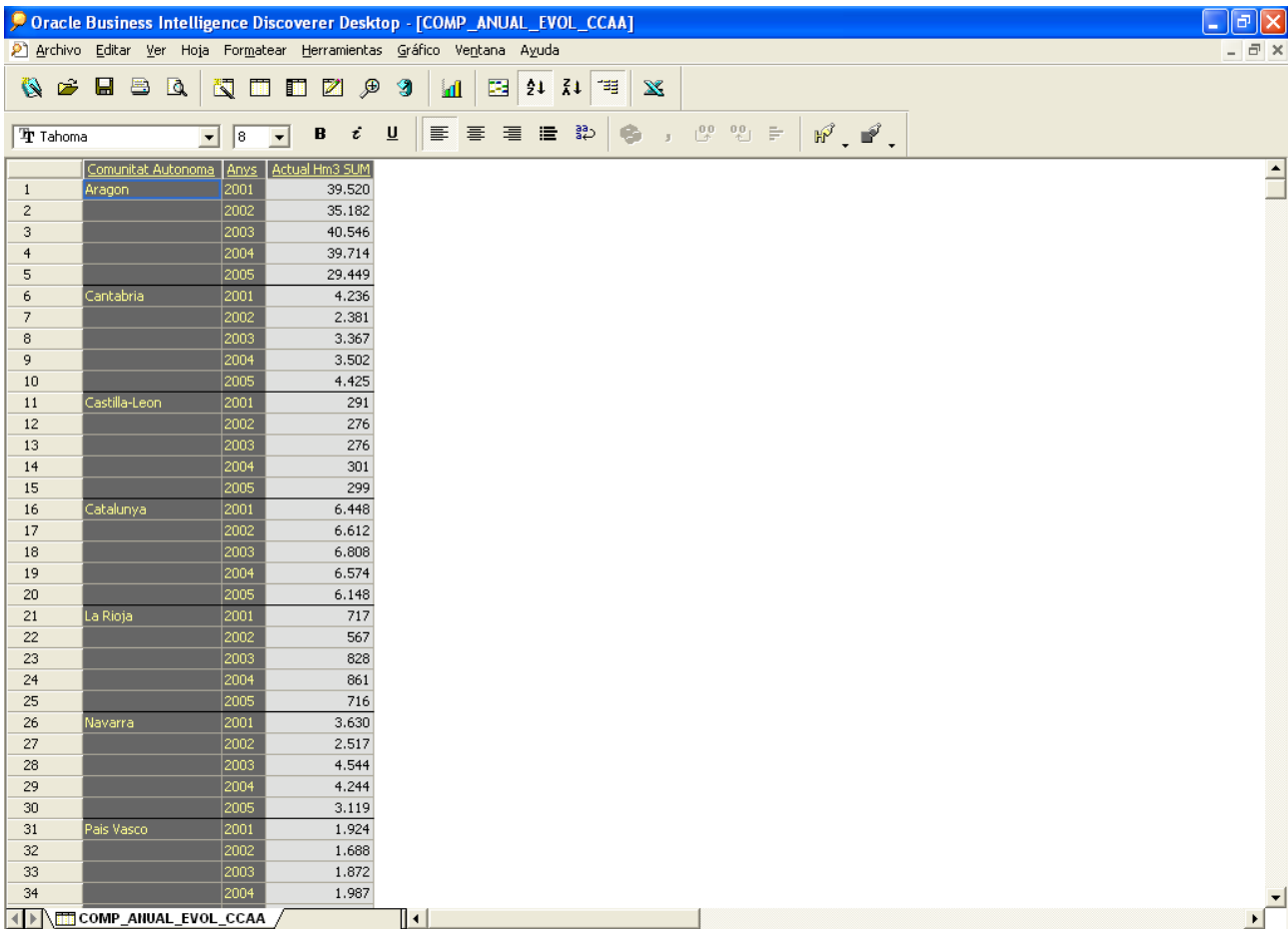
Elementos de Página:

	Comunitat Autònoma	Anys	Mes	Nom Mes	Total vs Actual SUM
1	Aragón	2001	01	Gener	747
2			02	Febrer	709
3			03	Març	553
4			04	Abril	670
5			05	Maig	533
6			06	Juny	653
7			07	Juliol	968
8			08	Agost	1.346
9			09	Setembre	1.646
10			10	Octubre	1.477
11			11	Novembre	1.329
12			12	Desembre	1.389
13		2002	01	Gener	1.365
14			02	Febrer	1.196
15			03	Març	1.096
16			04	Abril	1.052
17			05	Maig	894
18			06	Juny	1.018
19			07	Juliol	1.603
20			08	Agost	2.063
21			09	Setembre	2.033
22			10	Octubre	1.766
23			11	Novembre	1.388
24			12	Desembre	884
25		2003	01	Gener	768
26			02	Febrer	892
27			03	Març	769
28			04	Abril	576
29			05	Maig	562
30			06	Juny	610
31			07	Juliol	1.165
32			08	Agost	1.681

CAP_TOT_VS_ACT_CCAA_ANY_MES

Imatge 14. Informe 3

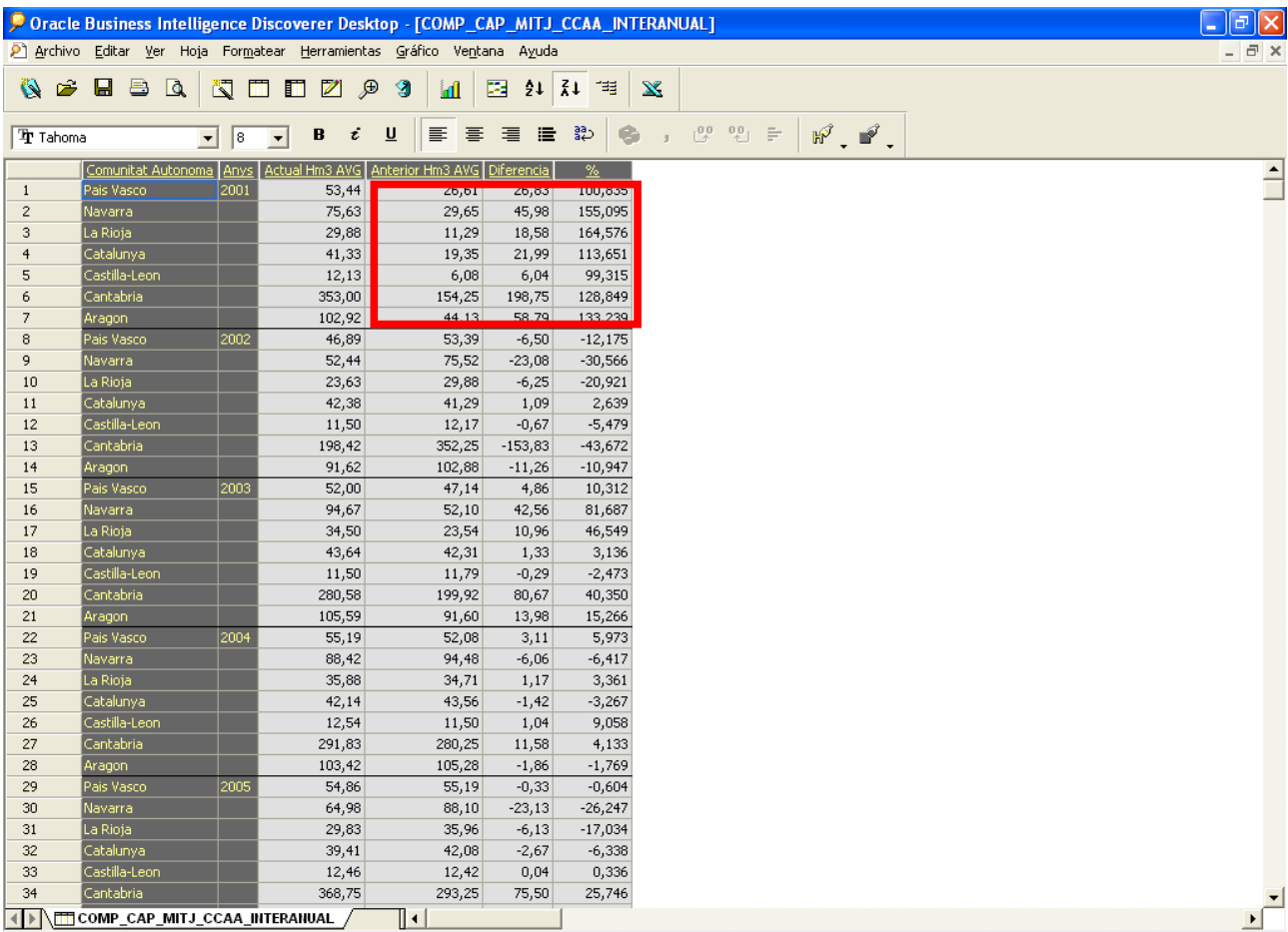
COMP ANUAL EVOL CCAA



	Comunitat Autònoma	Anys	Actual Hm3 SUM
1	Aragon	2001	39.520
2		2002	35.182
3		2003	40.546
4		2004	39.714
5		2005	29.449
6	Cantabria	2001	4.236
7		2002	2.381
8		2003	3.367
9		2004	3.502
10		2005	4.425
11	Castilla-Leon	2001	291
12		2002	276
13		2003	276
14		2004	301
15		2005	299
16	Catalunya	2001	6.448
17		2002	6.612
18		2003	6.808
19		2004	6.574
20		2005	6.148
21	La Rioja	2001	717
22		2002	567
23		2003	828
24		2004	861
25		2005	716
26	Navarra	2001	3.630
27		2002	2.517
28		2003	4.544
29		2004	4.244
30		2005	3.119
31	País Vasco	2001	1.924
32		2002	1.688
33		2003	1.872
34		2004	1.987

Imatge 15. Informe 4

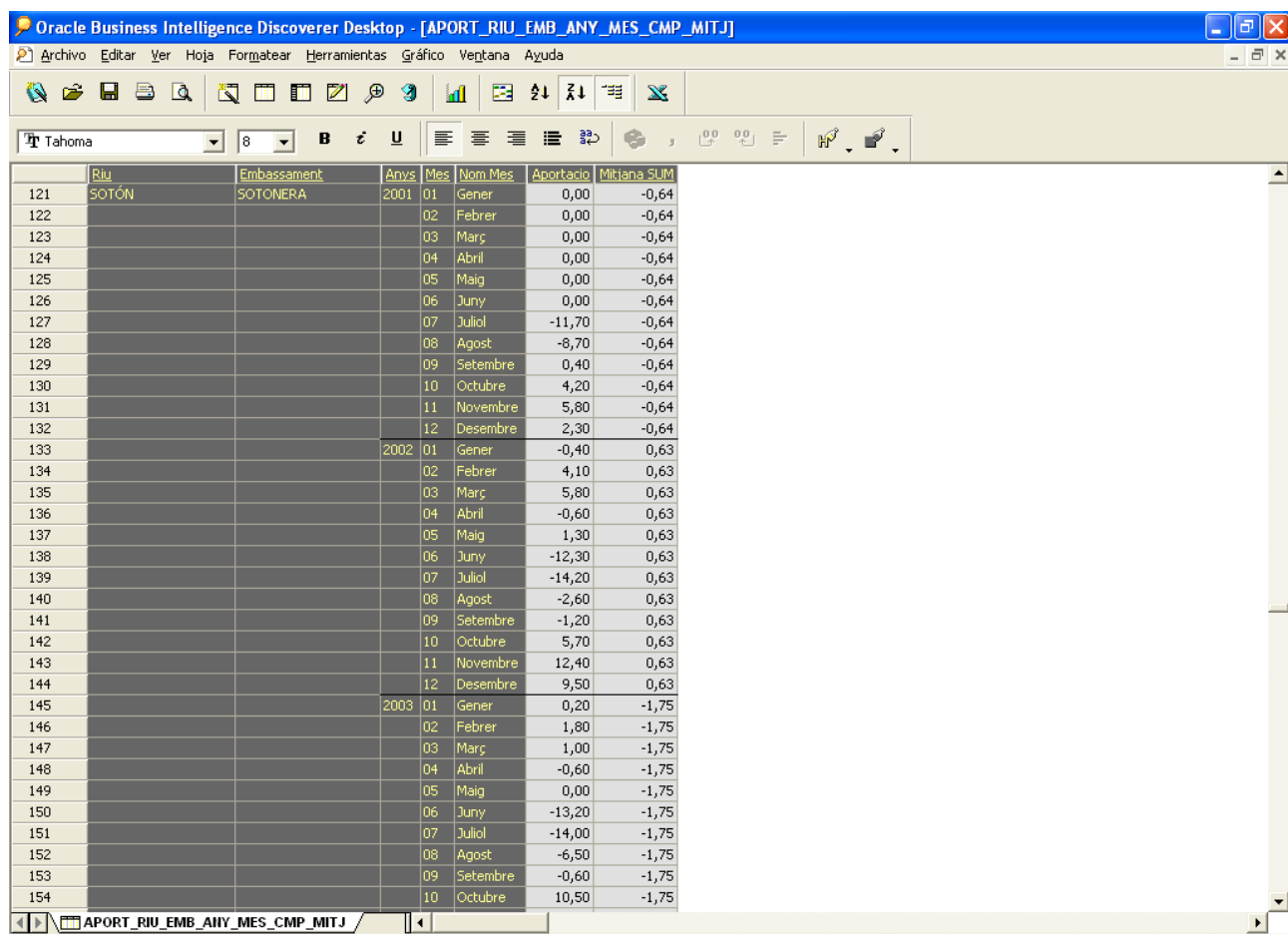
COMP_CAP_MITJ_CCAA_INTERANUAL



	Comunitat Autònoma	Anys	Actual Hm3 AVG	Anterior Hm3 AVG	Diferencia	%
1	Pais Vasco	2001	53,44	26,61	26,83	100,835
2	Navarra		75,63	29,65	45,98	155,095
3	La Rioja		29,88	11,29	18,58	164,576
4	Catalunya		41,33	19,35	21,99	113,651
5	Castilla-Leon		12,13	6,08	6,04	99,315
6	Cantabria		353,00	154,25	198,75	128,849
7	Aragon		102,92	44,13	58,79	133,239
8	Pais Vasco	2002	46,89	53,39	-6,50	-12,175
9	Navarra		52,44	75,52	-23,08	-30,566
10	La Rioja		23,63	29,88	-6,25	-20,921
11	Catalunya		42,38	41,29	1,09	2,639
12	Castilla-Leon		11,50	12,17	-0,67	-5,479
13	Cantabria		198,42	352,25	-153,83	-43,672
14	Aragon		91,62	102,88	-11,26	-10,947
15	Pais Vasco	2003	52,00	47,14	4,86	10,312
16	Navarra		94,67	52,10	42,56	81,687
17	La Rioja		34,50	23,54	10,96	46,549
18	Catalunya		43,64	42,31	1,33	3,136
19	Castilla-Leon		11,50	11,79	-0,29	-2,473
20	Cantabria		280,58	199,92	80,67	40,350
21	Aragon		105,59	91,60	13,98	15,266
22	Pais Vasco	2004	55,19	52,08	3,11	5,973
23	Navarra		88,42	94,48	-6,06	-6,417
24	La Rioja		35,88	34,71	1,17	3,361
25	Catalunya		42,14	43,56	-1,42	-3,267
26	Castilla-Leon		12,54	11,50	1,04	9,058
27	Cantabria		291,83	280,25	11,58	4,133
28	Aragon		103,42	105,28	-1,86	-1,769
29	Pais Vasco	2005	54,86	55,19	-0,33	-0,604
30	Navarra		64,98	88,10	-23,13	-26,247
31	La Rioja		29,83	35,96	-6,13	-17,034
32	Catalunya		39,41	42,08	-2,67	-6,338
33	Castilla-Leon		12,46	12,42	0,04	0,336
34	Cantabria		368,75	293,25	75,50	25,746

Imatge 16. Informe 5

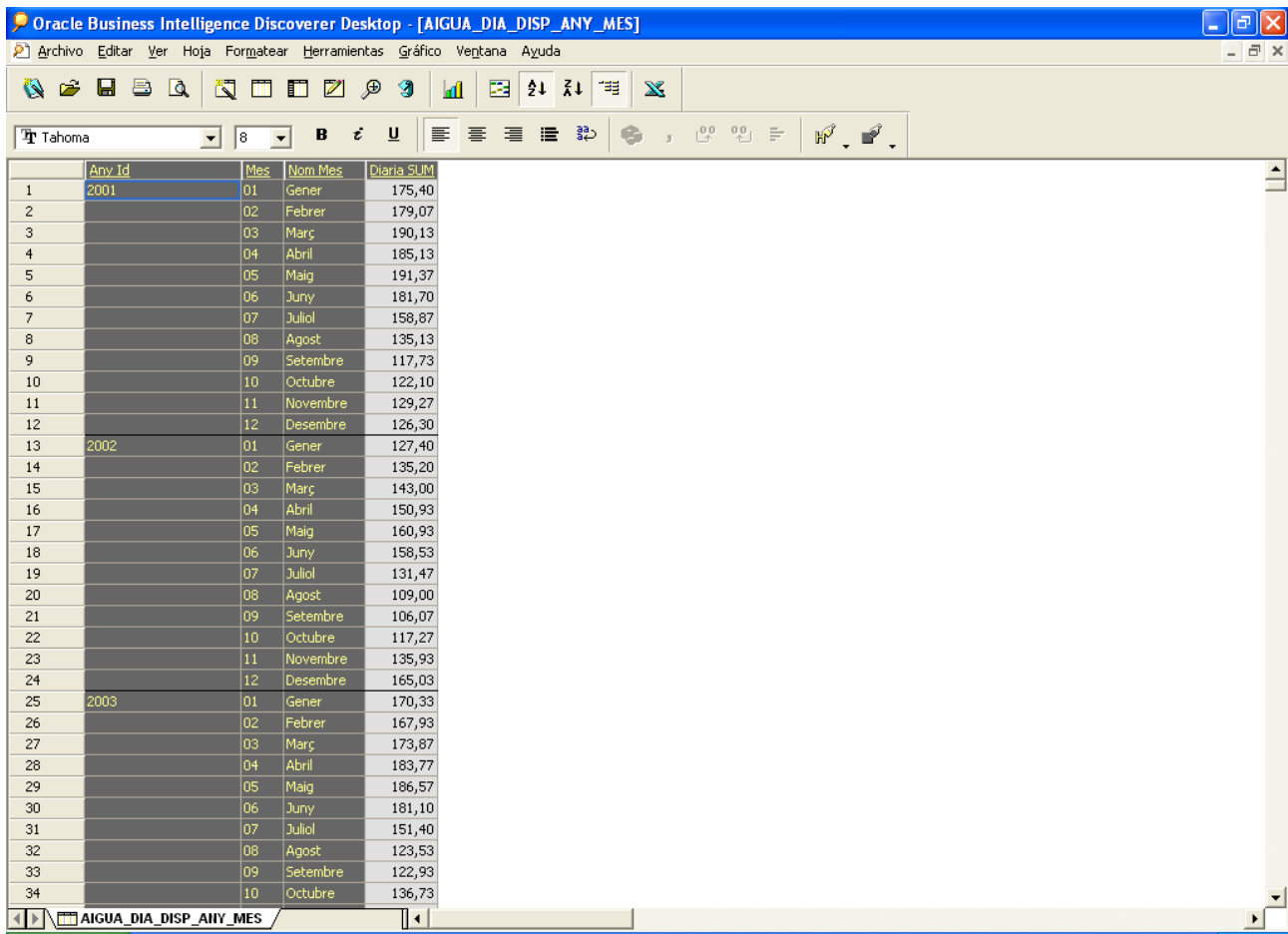
APORT RIU EMB ANY MES CMP MITJ



	Riu	Embassament	Anys	Mes	Nom Mes	Aportació	Mitjana SUM
121	SOTÓN	SOTONERA	2001	01	Gener	0,00	-0,64
122				02	Febrer	0,00	-0,64
123				03	Març	0,00	-0,64
124				04	Abril	0,00	-0,64
125				05	Maig	0,00	-0,64
126				06	Juny	0,00	-0,64
127				07	Juliol	-11,70	-0,64
128				08	Agost	-8,70	-0,64
129				09	Setembre	0,40	-0,64
130				10	Octubre	4,20	-0,64
131				11	Novembre	5,80	-0,64
132				12	Desembre	2,30	-0,64
133			2002	01	Gener	-0,40	0,63
134				02	Febrer	4,10	0,63
135				03	Març	5,80	0,63
136				04	Abril	-0,60	0,63
137				05	Maig	1,30	0,63
138				06	Juny	-12,30	0,63
139				07	Juliol	-14,20	0,63
140				08	Agost	-2,60	0,63
141				09	Setembre	-1,20	0,63
142				10	Octubre	5,70	0,63
143				11	Novembre	12,40	0,63
144				12	Desembre	9,50	0,63
145			2003	01	Gener	0,20	-1,75
146				02	Febrer	1,80	-1,75
147				03	Març	1,00	-1,75
148				04	Abril	-0,60	-1,75
149				05	Maig	0,00	-1,75
150				06	Juny	-13,20	-1,75
151				07	Juliol	-14,00	-1,75
152				08	Agost	-6,50	-1,75
153				09	Setembre	-0,60	-1,75
154				10	Octubre	10,50	-1,75

Imatge 17. Informe 6

AIGUA DIA DISP ANY MES

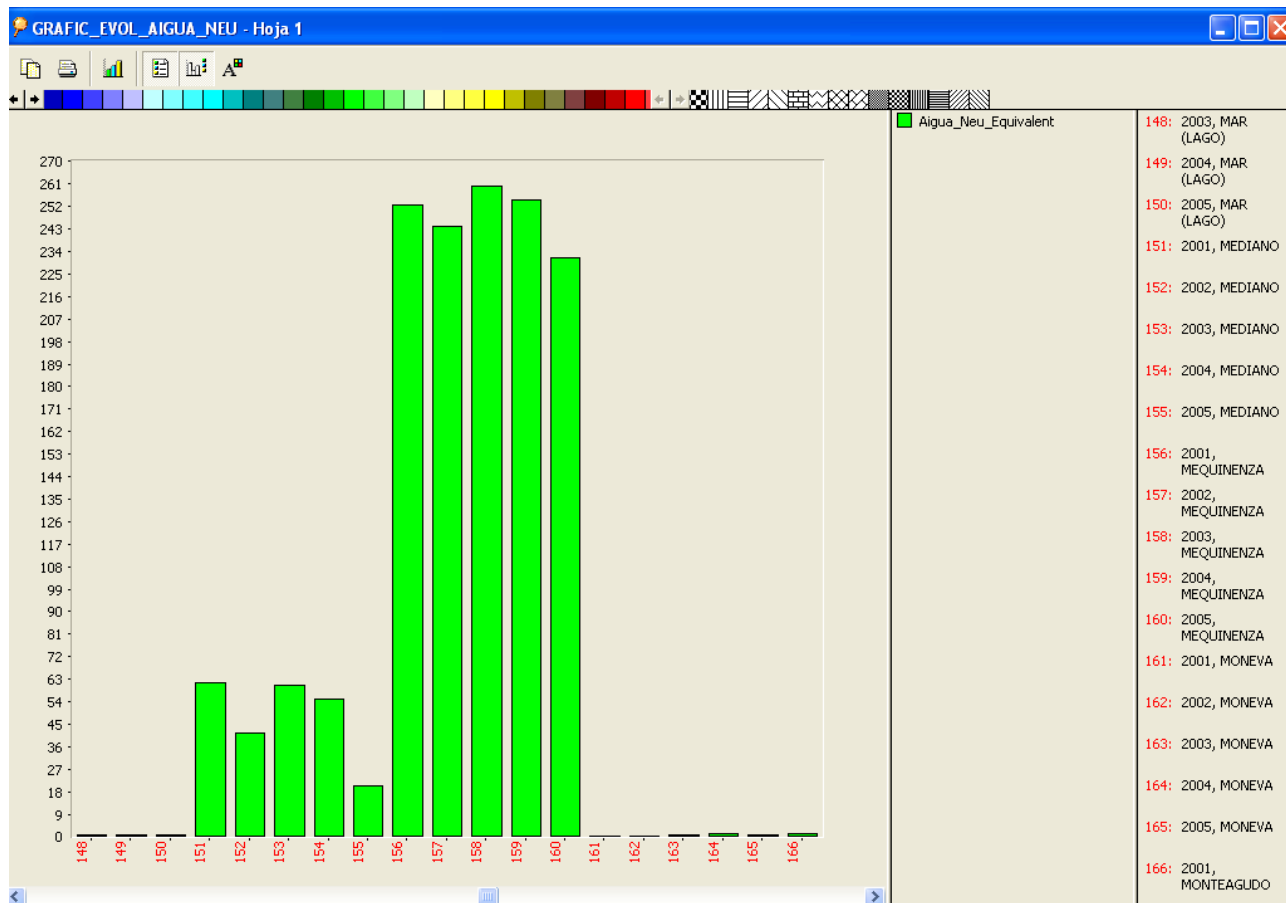


The screenshot shows the Oracle Business Intelligence Discoverer Desktop interface. The main window displays a table with the following data:

	Any Id	Mes	Nom Mes	Diaria SUM
1	2001	01	Gener	175,40
2		02	Febrer	179,07
3		03	Març	190,13
4		04	Abril	185,13
5		05	Maig	191,37
6		06	Juny	181,70
7		07	Juliol	158,87
8		08	Agost	135,13
9		09	Setembre	117,73
10		10	Octubre	122,10
11		11	Novembre	129,27
12		12	Desembre	126,30
13	2002	01	Gener	127,40
14		02	Febrer	135,20
15		03	Març	143,00
16		04	Abril	150,93
17		05	Maig	160,93
18		06	Juny	158,53
19		07	Juliol	131,47
20		08	Agost	109,00
21		09	Setembre	106,07
22		10	Octubre	117,27
23		11	Novembre	135,93
24		12	Desembre	165,03
25	2003	01	Gener	170,33
26		02	Febrer	167,93
27		03	Març	173,87
28		04	Abril	183,77
29		05	Maig	186,57
30		06	Juny	181,10
31		07	Juliol	151,40
32		08	Agost	123,53
33		09	Setembre	122,93
34		10	Octubre	136,73

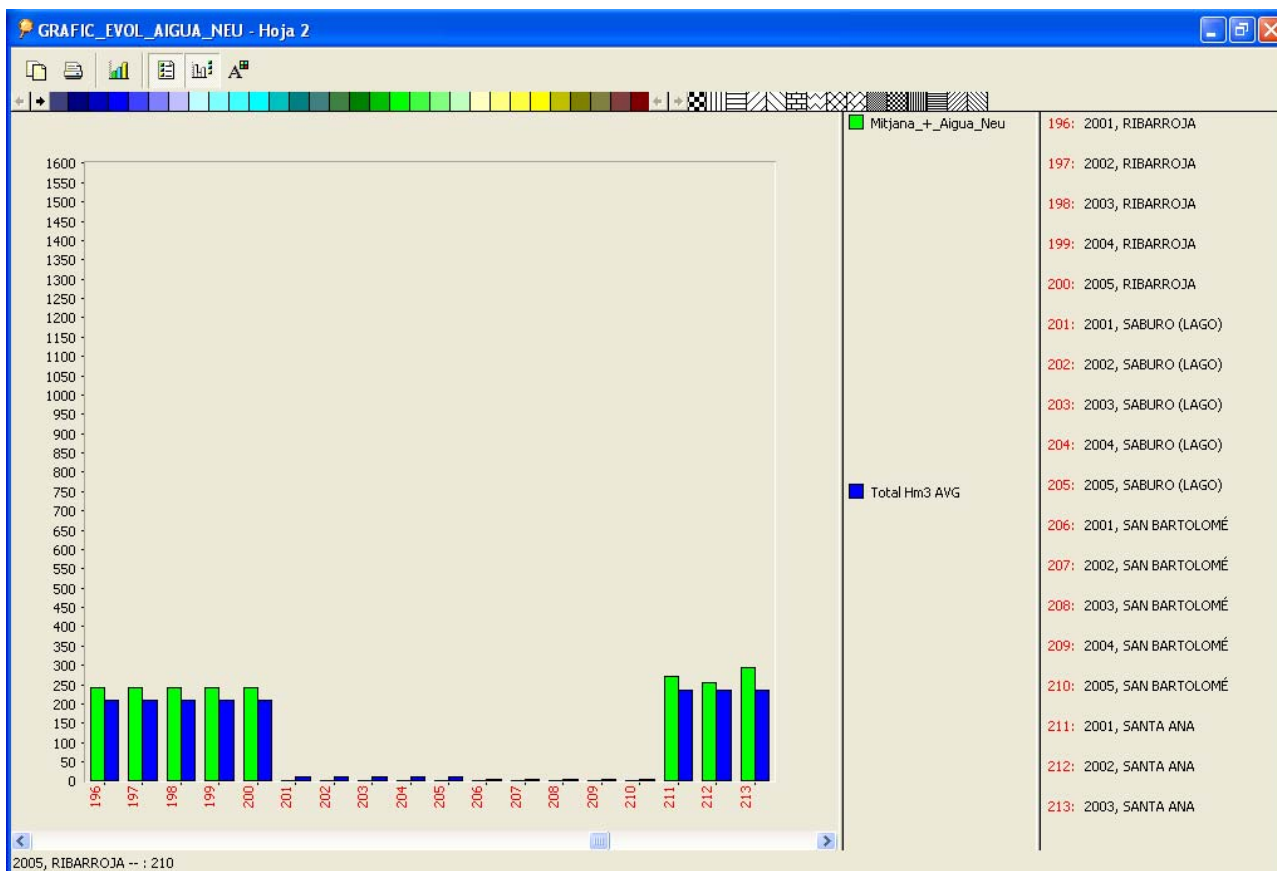
Imatge 18. Informe 7

GRAFIC EVOL AIGUA NEU



Imatge 19. Informe 8

Desbordaments



Imatge 20. Informe 9

En aquest gràfic s'observa clarament quan es poden produir casos de desbordament. Si la mitjana de la capacitat + l'aigua de Neu (barra en verd) es superior a la capacitat total (barra en blau) es produirà desbordament.

4. Conclusions

Bona part de la meua trajectòria professional ha anat lligada al món de les bases de dades i el reporting però sempre des de la vessant d'usuari final. El interès per aquest món va motivar que comencés aquest estudi.

Aquest projecte ha estat la peça perfecte per tancar aquest cicle d'aprenentatge. M'ha permès estructurar i donar coherència a un seguit de coneixements adquirits tant en el món laboral com durant aquests anys de carrera.

Com a punt positiu puc dir que aquest projecte m'ha ajudat a veure el bosc sencer i no només un grup aïllat d'arbres. He pogut veure l'origen, les necessitats d'un tercer i el seguit de passos entremitjos per arribar a un producte final.

Com a punt negatiu he de reconèixer que la part de generació de documentació podia haver estat força millor. El meu perfil és més de "fer coses" que "d'explicar com s'han fet". Es un punt ha millorar sobretot si em vull dedicar professionalment a l'àrea de la Business Intelligence, com seria el meu cas.

5. Línies d'evolució futures

Com a possibles millores de cara al futur es poden exposar les següents:

Al projecte es va partir d'un seguit de dades històriques però no s'ha plantejat com fer la captura de noves dades en el futur. S'hauria de crear una plantilla d'entrada de dades amb l'estructura adequada que facilités la càrrega a la base de dades.

Pensant en l'expansió del negoci es pot intentar ampliar el producte per incloure més confederacions hidrogràfiques. Això implicaria unes mínimes modificacions tant a la base de dades com als informes per tal d'incorporar una nova dimensió, **la confederació hidrogràfica**.

En posteriors fases es podria "internacionalitzar" el producte i incorporar nous països.

-Crear informes addicionals.

-Introducció d'eines de data – Mining.

-Intentar recopilar informació històrica anterior al 2001.

6. Glossari

Business Intelligence: Disciplina que persegueix l'explotació de dades estructurades per recolzar la presa de decisions.

ETL: Extraction, Transformation & Loading: Procés mitjançant el qual s'extreuen dades d'un o més d'un origen de dades, es manipulen per tal d'adaptar-les a l'estructura desitjada i finalment es carreguen al sistema de destí.

Magatzem de dades: Tipologia de base de dades estructurada en cubs multidimensionals amb l'objectiu d'obtenir informació de forma eficient.

Dimensions: Són el eixos d'anàlisi de la nostra informació. Són elements fàcils d'identificar a la vida real.

Atributs: Característiques pròpies de les dimensions.

Indicadors: són les dades numèriques que s'analitzen.

7. Bibliografia

Rius Gavídia, A., Serra Vizern, M.

Magatzems de dades i models multidimensionals

UOC, Material assignatura

Ralph Kimball and Margy Ross
The Data Warehouse Toolkit (2nd Edition)